SEMANTIC PROCESS ANALYSIS, CONTEXT-AWARE INFORMATION RETRIEVAL, AND SENTIMENT ANALYSIS FOR SUPPORTING TRANSPORTATION PROJECT ENVIRONMENTAL REVIEW

BY

XUAN LV

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Civil Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Associate Professor Nora El-Gohary, Chair and Director of Research
Professor Chengxiang Zhai
Professor Khaled El-Rayes
Associate Professor Liang Y. Liu
Associate Professor Mani Golparvar-Fard

ABSTRACT

According to the National Environmental Policy Act (NEPA), transportation projects are required to go through an environmental review process to evaluate their impact on the environment. However, the Transportation Project Environmental Review process (TPER), has long been "criticized for resulting in frequent delays in the development of important projects designed to improve the safety and operating conditions of a region's transportation system" (FHWA 2013); the time to complete the environmental review process for large-scale transportation projects nearly tripled since the 1970s (Clark and Canter 1997; Barberio et al. 2008a; Venner Consulting et al. 2012). Based on a number of studies (e.g., Mallett and Luther 2011; Cambridge Systematics, Inc. 2011; Keck et al. 2010; FHWA 2016) conducted to identify the constraints for accelerating the TPER process, three primary causes of process inefficiencies were identified: (1) NEPA and transportation project planning processes are not streamlined; (2) transportation practitioners have limited ability to find the right information, at the right time to support mission-critical analyses (Spy Pond Parteners et al. 2009): and (3) there is late identification of stakeholder concerns and support levels.

Towards addressing these three problems, this thesis aims to enhance the efficiency of the TPER process through (1) discovering the practices that should be implemented to integrate the NEPA process into the transportation planning process in a manner to ensure both the efficiency of project development and compliance with NEPA; (2) developing context-aware information retrieval methods to support the search and retrieval of relevant textual information in the TPER domain; and (3) developing stakeholder opinion mining methods to identify potential concerns and stakeholder support levels early in the project development process.

Accordingly, the thesis includes eight primary research tasks: (1) conducting a comprehensive literature review; (2) analyzing existing processes and identifying successful integration practices for integrating NEPA into transportation planning processes for large-scale highway projects in Illinois; (3) developing a semantic annotation method and algorithm for supporting context-aware information retrieval in the TPER domain; (4) developing a semantic, context-aware information retrieval method and algorithm for retrieving relevant information for supporting the TPER process; (5) developing a stakeholder opinion extraction method and algorithm for automatically extracting subject, concern, and opinion expressions from stakeholder comments on large-scale highway projects to support aspect-level stakeholder opinion mining in the TPER domain; (6) developing a stakeholder opinion classification method and algorithm for classifying the extracted subject, concern, and opinion expressions to support aspect-level opinion mining in the TPER domain; (7) developing a sentence-level opinion mining method and algorithm for classifying comment sentences on large-scale highway projects; and (8) conducting case studies to analyze the differences and similarities among different stakeholder groups in terms of concerns and support levels.

All proposed methods and algorithms were tested and evaluated, and the results of these evaluations are presented in the thesis. The thesis also discusses the limitations and recommendations for future research.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincerest gratitude to my advisor Prof. Nora El-Gohary for her continuous guidance and support during my Ph.D. study. I would also like to deeply thank my Doctoral Committee Members – Prof. Chengxiang Zhai, Prof. Khaled El-Rayes, Prof. Liang Y. Liu, and Prof. Mani Golparvar-Fard – for their insightful comments and advice.

I would like to give special thanks to my wife, Lu Zhang, for her love, understanding, patience, and unconditional support throughout my Ph.D. study. She has been incredibly supportive and helpful, especially during my thesis writing time. Special thanks also go to my parents and the rest of my family for their love, encouragement, and support in my life.

I thank my former groupmates, Prof. Jiansong Zhang and Prof. Lu Zhang, and my current groupmates – Peng Zhou, Marwan Ammar, Kadir Amasyali, Lufan Wang, Kaijian Liu, Ruichuan Zhang, and Peter Liu – for their stimulating discussions and inspiring presentations, and for all the fun we had in the past five years. I also thank all the rest of my friends and colleagues at the University of Illinois at Urbana-Champaign.

Finally, I gratefully acknowledge the funding support from the Illinois Center of Transportation, the Illinois Department of Transportation, and the Strategic Research Initiatives (SRI) Program by the College of Engineering, University of Illinois at Urbana-Champaign.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Motivation and Overview

According to the National Environmental Policy Act (NEPA), transportation projects are required to go through an environmental review process to evaluate their impact on the environment. The environmental review process not only affects transportation decision making by taking environmental concerns into account, but also affects the project development process in terms of time and cost. According to a study conducted on the timeliness of the environmental review process (Venner Consulting et al. 2012), the environmental review process consumes nearly 30% of the total project development time on average, and a longer review time is correlated with a longer project development time. The Transportation Project Environmental Review process (TPER), which requires the collaboration of a number of stakeholders and the collection and communication of a large amount of textual information, has long been "criticized for resulting in frequent delays in the development of important projects designed to improve the safety and operating conditions of a region's transportation system" (FHWA 2013); the time to complete the environmental review process for large-scale transportation projects nearly tripled since the 1970s (Clark and Canter 1997; Barberio et al. 2008a; Venner Consulting et al. 2012).

There have been many administrative and legislative efforts (USGPO 1998; USGPO 2007; USGPO 2013) to expedite the environmental review process and a number of studies (e.g., Mallett and Luther 2011; Cambridge Systematics, Inc. 2011; Keck et al. 2010; FHWA 2016) were conducted to identify the primary causes of inefficiencies in the TPER process and pinpoint opportunities for improvement. Based on these studies, three primary causes of process inefficiencies (and consequently longer TPER process durations) were identified. First, NEPA and transportation project planning processes are not streamlined. When NEPA and transportation project planning processes are not streamlined, the NEPA process may lead to duplication of work and project delays (Keck et al. 2010). Second, substantial gaps exist in the ability of transportation practitioners to find the right information, at the right time, for the task at hand (Spy Pond Parteners et al. 2009). During the transportation project planning process, duplication of efforts can be avoided by learning from previous cases, i.e., environmental reviews conducted for similar types of projects that potentially impact similar environmental resources. This requires searching for and finding such relevant environmental reviews and associated documents. However, "finding the right information to support mission-critical analysis and decision making is often difficult" (TRB 2014); it is estimated that "80-90% of information is unstructured and that an agency's employees may spend up to 35% of their time looking for information" (Cambridge Systematics, Inc. 2013). Third, projects suffer from late identification of stakeholder concerns and support levels. Stakeholder typically have concerns about environmental, cultural, and socioeconomic issues. Understanding and addressing these concerns – and ensuring that the stakeholder are supportive of the project – at the early planning stage is crucial for project success. Late identification of stakeholder concerns and support levels could lead to design changes, reevaluation of already completed studies, and additional public consultation and stakeholder involvement efforts – all which could cause serious project delays and cost overruns.

Towards addressing the aforementioned three problems, this thesis aims to enhance the efficiency of the TPER process through (1) discovering the practices that should be implemented to integrate the NEPA process into the transportation planning processes in a manner to ensure both the efficiency of project development and compliance with NEPA; (2) developing context-aware information retrieval methods to support the search and retrieval of relevant textual information in the TPER domain; and (3) developing stakeholder opinion mining methods to identify potential concerns and stakeholder support levels early in the project development process. Accordingly, the thesis includes eight primary research tasks: (1) conducting a comprehensive literature review; (2) analyzing existing processes and identifying successful integration practices for integrating NEPA into transportation planning processes for large-scale highway projects in Illinois; (3) developing a semantic annotation method and algorithm for supporting context-aware information retrieval in the TPER domain; (4) developing a semantic, context-aware information retrieval method and algorithm for retrieving relevant information for supporting the TPER process; (5) developing a stakeholder opinion extraction method and algorithm for automatically extracting subject, concern, and opinion expressions from stakeholder comments on large-scale highway projects to support aspect-level stakeholder opinion mining in the TPER domain; (6) developing a stakeholder opinion classification method and algorithm for classifying the extracted subject, concern, and opinion expressions to support aspect-level opinion mining in the TPER domain; (7) developing a sentence-level opinion mining method and algorithm for classifying comment sentences on large-scale highway projects; and (8) conducting case studies to analyze the differences and similarities among different stakeholder groups in terms of concerns and support levels.

1.2 State of the Art and Knowledge Gaps

1.2.1 State of the Art and Knowledge Gaps in NEPA and Transportation Planning Integration

The federal government has developed several guidance documents for integrating NEPA into transportation project planning processes. Section 1309 of the Transportation Equity Act for the 21st Century (TEA-21) initiated the federal guidance for integrating the NEPA process into the state department of transportation (DOT) and metropolitan planning organization (MPO) planning processes in 1998 (USGPO 1998); it mandated the development and implementation of a coordinated environmental review process especially for projects that require the preparation of an environmental impact statement (EIS). In 2007, Section 6002 of the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU) established a new environmental review process for transportation projects that require an EIS in order to promote efficient project management and enhanced interagency coordination (USGPO 2007). The most up-to-date federal guidance is provided in Section 1301-1323 of the Moving Ahead for Progress in the 21st Century Act (MAP-21) (USGPO 2013). These three sections introduce programmatic approaches to promote greater linkages between the planning process and the environmental review process, and establish frameworks for setting deadlines for decision making during the environmental review process considering conflict resolution and penalties for agencies that fail to make a decision (USGPO 2013).

In response to the federal guidance, a number of states have conducted extensive research studies on how to integrate the NEPA process into their transportation project planning processes, and developed detailed and formal guidelines on how to implement and evaluate the integrated process. For example, the Colorado Department of Transportation (CDOT) developed the Strategic Transportation, Environmental, and Planning Process for Urban Places (STEP-UP), which included the use of a Geographic Information System (GIS)-based tool for identifying and assessing the environmental impacts and a methodology for conducting regional cumulative effect assessment (FHWA 2007a; MacDonald and Lidov 2007). The Florida Department of Transportation (FDOT) developed the Efficient Transportation Decision-Making (ETDM) process, which utilized the Environmental Screening Tool (EST), an internet-accessible interactive database for documenting project changes, evaluating impacts, and communicating project details to agencies and the public (FDOT 2006; FHWA 2007b). The Indiana Department of Transportation (INDOT) developed a streamlined procedure to eliminate the duplication of activities between the planning studies and the subsequent environmental analyses carried out under NEPA (FHWA 2007c; INDOT and FHWA 2007). The Maine Department of Transportation (MaineDOT) developed the Maine's Integrated Transportation Decision Making (ITD) process for projects that require an EIS or EA (FHWA 2002; FHWA 2007d).

Although a number of studies have been conducted in other states on integrating NEPA into their transportation planning processes, three primary knowledge gaps are identified. First, there is lack of integration efforts that focus on integrating NEPA into transportation planning at both the system and the corridor levels. Previous integration efforts either integrated NEPA with system-level planning (statewide level or metropolitan level) or with corridor-level planning. For example, Colorado's STEP-UP process incorporated environmental review into the North Front Range MPO's regional planning process (MacDonald and Lidov 2005), and Indiana's streamlined EIS procedure integrated corridor-level planning and NEPA studies in one decision-making process (INDOT and FHWA 2007). Second, there is lack of implementation detail on how to conduct environmental analysis during the planning process. For example, Maine's ITD process did not provide implementation details on how to conduct environmental analysis during the different phases of planning (FHWA 2007d) and Florida's EDTM process did not provide implementation details on how to incorporate the findings of planning-level environmental analysis into future NEPA decision making (FDOT 2006). Providing such detail, finding the right level of detail, and offering detail that is context-specific is essential for successful integration; to incorporate information from the planning process into the subsequent NEPA process, the planning-level environmental analysis should not only be accurate and up-to-date, but should also contain the level of detail that is compliant with NEPA requirements (Barberio et al. 2008b). The implementation detail also depends on the implementation context, in terms of the characteristics of the environmental issues, the current conditions, and the availability of resources (e.g., data, analysis tool). Third, there is lack of standardized/formalized performance measures to evaluate the implementation of integrated planning and NEPA processes. Developing standardized/formalized performance measures is important to help better demonstrate the qualitative and quantitative improvements in terms of project delivery and compliance with NEPA, which may further promote the implementation of process integration efforts. Existing integration efforts either did not develop any performance measure, such as Colorado's STEP-UP process (FHWA 2007a), or lacked performance measures for important planning studies, such as Florida's EDTM process which did not include performance measures on evaluating corridor/feasibility studies (FDOT 2005).

1.2.2 State of the Art and Knowledge Gaps in Information Retrieval

In recent years, a number of important research efforts have been conducted for improving information retrieval in the construction domain. For example, Soibelman et al. (2007) combined a vector space model with document classification information to retrieve documents related to a project model object, and developed a domain-specific thesaurus to improve the retrieval of construction product information from the internet. Lin and Soibelman (2009) proposed a domain-specific search engine for architectural/engineering/construction (AEC) online products, which incorporated domain knowledge about products through query expansion and extended Boolean model retrieval. McGibbney and Kumar (2011) developed a web-based information search and retrieval framework that utilized a domain ontology to facilitate the retrieval of energy performance building regulations, which integrated ontology-enhanced query refinement. Demian and Balastoukas (2012) investigated the effects of granularity and context when measuring relevance and visualizing results for retrieving building design and construction content, and found that users performed better and were more satisfied when the search results were displayed with their context information in terms of the related discipline, building components, and subcomponent objects. Fan et al. (2015) implemented three machine learning algorithms to enhance the retrieval results through user feedback, and utilized a project-specific term dictionary and dependency grammar information to facilitate feature selection.

Outside of the construction domain, but in the engineering domain, a number of important information retrieval research efforts have also been conducted. For example, Liu et al. (2006) proposed a framework to retrieve specific engineering document fragments with precise, complex queries, which integrates five modules including document navigation system, fragment classification, fragment extraction, document mark-up, and document structure, and demonstrated its advantages over general search engines when retrieving document fragments. Hahm et al. (2014) developed an ontology-based, personalized query expansion method to retrieve engineering documents with less semantic ambiguity and more focus on personalized information needs, which generates a user's profile from the domain ontology, and refines it through relation weighting. Hahm et al. (2015) proposed a document ranking approach that incorporates relationships among terms in the relevance assessment process based on a domain ontology, which represents the semantics of a document through a document semantic network and considers both user interests and searching intent through relation-based weighting.

Despite the importance of the aforementioned information retrieval research efforts, there still exist many challenges in developing information retrieval methods that can efficiently retrieve relevant information for transportation project decision making: most of the existing information retrieval research efforts and systems in the construction domain are limited in their context-awareness. Such limitation could be attributed to two main reasons. First, many of the existing information retrieval efforts build on keyword-based content representation and query processing techniques, which provide limited capabilities for incorporating content semantics and contextual information into the retrieval process (Fernandez et al. 2011). Keyword-based information retrieval methods can, therefore, be very ineffective when handling context-sensitive tasks, such as searching for environmental review studies of similar projects in terms of project type, location, and resources affected. Second, existing semantic-based information retrieval efforts are limited in both context representation and context-based retrieval. On one hand, these efforts are limited in their formal context representation – they lack an explicit, domain-specific representation of the concept of context. Limited context representation limits their capability to recognize domain-specific contextual information in both the users' queries and the documents. On the other hand, these efforts are limited in their domain-specific context-based information retrieval – they lack support for semantic query processing and semantic document ranking based on multiple context dimensions – including user context, process context, and resource context. Limited context-based retrieval makes these systems ineffective in supporting domain-specific decision making, because real-life transportation project decision making scenarios are related to those contextual dimensions. In addition to these two limitations, most of the existing information retrieval efforts in the construction domain have not compared their methods to other state-of-the-art information retrieval methods (in other domains) in terms of retrieval performance.

1.2.3 State of the Art and Knowledge Gaps in Stakeholder Opinion Mining

In recent years, a number of research studies have been conducted on applying texting mining techniques in the construction domain. For example, Williams and Gong (2013) applied data mining classification algorithms to predict of the level of cost overrun based on text descriptions of a project's characteristics and numerical data. Alsubaey et al. (2015) proposed a Naïve Bayes text mining approach to identify early warnings of project failures based on critical management documents such as minutes of meetings. Nik-Bakht and El-Diraby (2016) combines community detection in social networks with information retrieval methods to detect and label communities of project followers and cores of interest in the network of urban infrastructure project stakeholders.

In the computer science domain, relevant research on stakeholder opinion mining has also been extensively studied in the literature. Qu et al. (2010) utilized a constrained ridge regression algorithm to predict a users' numeric rating of products based on the user's product review text. They proposed a bag-of-opinions representation of review text that outperformed the traditional unigram and n-gram representations. To produce good quality summary of opinions, Zhai et al. (2011) proposed a semi-supervised learning approach for clustering or grouping synonym features from users' reviews. They utilized lexical characteristics to automatically identify some labeled examples and applied an expectation and maximization (EM) algorithm for training. The proposed semi-supervised approach outperformed the state-of-the-art unsupervised approach by a large margin. Anjaria and Guddeti (2014) proposed a hybrid approach of extracting opinion using direct and indirect features of Twitter data based on a number of supervised classifiers. They conducted two case studies of using twitter to predict election results and concluded the conditions of failure and success.

However, the current research efforts in the area of stakeholder opinion mining are limited in supporting the identification of stakeholder concerns and support levels early in the project development process, because of three reasons. First, most of the opinion mining research efforts focused on stakeholder opinions on products or services, which are different from stakeholder opinions on large-scale transportation projects in terms of opinions and concerns expressed, and the linguistic patterns displayed. Second, most of such efforts focused on one stakeholder group (e.g., users of the product/service), while there are multiple stakeholder groups (e.g., resource agencies, residents, land owners) in identifying stakeholder concerns and support levels for transportation projects. Third, most of such efforts focused on analyzing sentiments expressed by the comments, and have limited ability to identify concerns from stakeholder comments.

1.3 Problem Statement

There is an increasing demand to improve the efficiency of the current TPER process. Three primary causes of process inefficiencies were identified based on previous studies: (1) NEPA and transportation project planning processes are not streamlined; (2) substantial gaps exist in the ability of transportation practitioners to find the right information, at the right time, for the task at hand; and (3) there is late identification of stakeholder concerns and support levels. For streamlining NEPA and transportation project planning processes, previous integration efforts are limited in three main ways: (1) three is lack of efforts that focus on integrating NEPA into transportation planning at both the system and the corridor levels; (2) there is lack of implementation detail on how to conduct environmental analysis during the planning process; and (3) there is lack of standardized/formalized performance measures to evaluate the implementation of integrated planning and NEPA processes. For improving information retrieval in the TPER domain, previous information retrieval efforts in the construction domain are limited in three main ways: (1) they mostly build on keyword-based content representation and query processing techniques, which provide limited capabilities for incorporating content semantics and contextual information into the retrieval process; (2) they are limited in their formal context representation and context-based retrieval, which provide limited capabilities to recognize domain-specific contextual information and to support context-based information retrieval on multiple context dimensions; and (3) they have not been compared with state-of-the-art information retrieval systems outside of the construction domain in terms of retrieval performance. For facilitating the identification of stakeholder concerns and support early in the project development process, previous efforts on stakeholder opinion mining are limited in three main ways: (1) they focused on stakeholder opinions on products or services, which are different from stakeholder opinions on transportation projects in terms of opinions and concerns expressed, and the linguistic patterns displayed; (2) they focused on one stakeholder group (e.g., users of the product/service), while there are multiple stakeholder groups (e.g., resource agencies, residents, land owners) in identifying stakeholder concerns and support for transportation projects; and (3) they focused on analyzing sentiments expressed by the comments, and have limited ability to identify concerns from stakeholder comments.

1.4 Research Objectives and Questions

This thesis aims to enhance the efficiency of the TPER process through (1) discovering the practices that should be implemented to integrate the NEPA process into state DOT and MPO planning processes for large-scale highway projects in Illinois in a manner to ensure both the efficiency of project development and compliance with NEPA; (2) developing context-aware information retrieval methods to support the search and retrieval of relevant textual information in the TPER domain; and (3) developing stakeholder opinion mining methods to identify stakeholder concerns and support levels for large-scale highway projects early in the project development process. Accordingly, seven specific objectives and outcomes are defined.

Objective 1: Discover the integration practices for integrating the NEPA process into the state DOT and MPO planning processes for large-scale highway projects in the state of Illinois.

Research Questions: What are the potential integration practices? What are the suitable integration practices for the state of Illinois that should be selected from these potential integration practices? How to incorporate NEPA with transportation planning at both the system level and the corridor level? How to implement the selected integration practices in Illinois? What are the performance measures for evaluating the implementation of the integrated process?

Outcome: (a) Identifying potential integration practices, (b) identifying state-suitable integration practices for the state of Illinois, (c) developing a process model for the integrated Illinois Department of Transportation (IDOT)-MPO-NEPA planning process, and (d) providing well-defined guidance on the implementation and evaluation of the integrated process.

Objective 2: Develop a semantic annotation method for automatically annotating textual documents with TPER-domain-specific semantic concepts for supporting context-aware information retrieval in the TPER domain.

Research Questions: How to automatically annotate textual documents with semantic concepts that are relevant to the TPER domain? What are these semantic concepts and how to best model them? How is the performance of shallow semantic annotation methods compared with deep semantic annotation methods?

Outcome: A semantic annotation method and algorithm that automatically annotates documents with TPER-domain-specific semantic concepts.

Objective 3: Develop a semantic, context-aware information retrieval method for retrieving relevant information in the TPER domain.

Research Questions: How to conduct semantic query processing to automatically extract context information from user queries? How to conduct semantic document ranking to rank the retrieved documents based on the context information? How is the performance of vector-space-model-based methods compared with statistical-language-model-based methods?

Outcome: A semantic context-aware information retrieval method and algorithm for retrieving relevant information in the TPER domain.

Objective 4: Develop a stakeholder opinion extraction method for automatically extracting subject, concern, and opinion expressions from stakeholder comments on large-scale highway projects to support aspect-level stakeholder opinion mining in the TPER domain.

Research Questions: How to automatically extract subject, concern, and opinion expressions from stakeholder comments? What are the machine learning algorithms to use for the extraction? What are the best features to use for the extraction?

Outcome: A stakeholder opinion extraction method and algorithm that automatically extracts subject, concern, and opinion expressions from stakeholder comments on large-scale highway projects to support aspect-level stakeholder opinion mining in the TPER domain.

Objective 5: Develop a stakeholder opinion classification method for classifying extracted subject, concern, and opinion expressions (opinion tuples) into concern and sentiment categories to support aspect-level stakeholder opinion mining in the TPER domain.

Research Questions: How to automatically classify opinion tuples from stakeholder comments into concern and sentiment categories? How to develop an unsupervised method for classifying opinion tuples into concern and sentiment categories, to save manual effort? How is the classification performance of the unsupervised method compared with existing supervised methods?

Outcome: A stakeholder opinion classification method and algorithm that classifies extracted opinion tuples into concern and sentiment categories to support aspect-level stakeholder opinion mining in the TPER domain.

Objective 6: Develop a sentence-level stakeholder opinion mining method for automatically classifying sentences from stakeholder comments on large-scale highway projects into concern and sentiment categories. Compared to the tuple-based method (Objectives 4 and 5), the sentence-level method offers an alternative approach when a sentence-level analysis is sufficient.

Research Questions: How to automatically classify sentences from stakeholder comments into concern and sentiment categories? How to develop an unsupervised method for this classification problem, to save manual effort? How is the classification performance of the unsupervised method compared with the supervised approach?

Outcome: A sentence-level stakeholder opinion mining method and algorithm for classifying sentences from stakeholder comments on large-scale highway projects into concern and sentiment categories.

Objective 7: Analyze stakeholder comments from a set of case study projects to gain a better understanding of stakeholder opinions, and how they could be similar or different across different stakeholder groups.

Research Questions: What are the support levels of the project stakeholders? What are the concerns of the stakeholders? What are the negative concerns of the stakeholders? What are the similarities and differences – in support levels, concerns, and negative concerns – across the different stakeholder groups?

Outcome: A better understanding of stakeholder opinions, and their similarities and differences among different stakeholder groups, based on a set of case study projects.

1.5 Research Methodology and Tasks

The research methodology includes eight primary research tasks, as summarized in Figure 1.1. A more detailed explanation of the methodology of each task is presented in the following subsections.

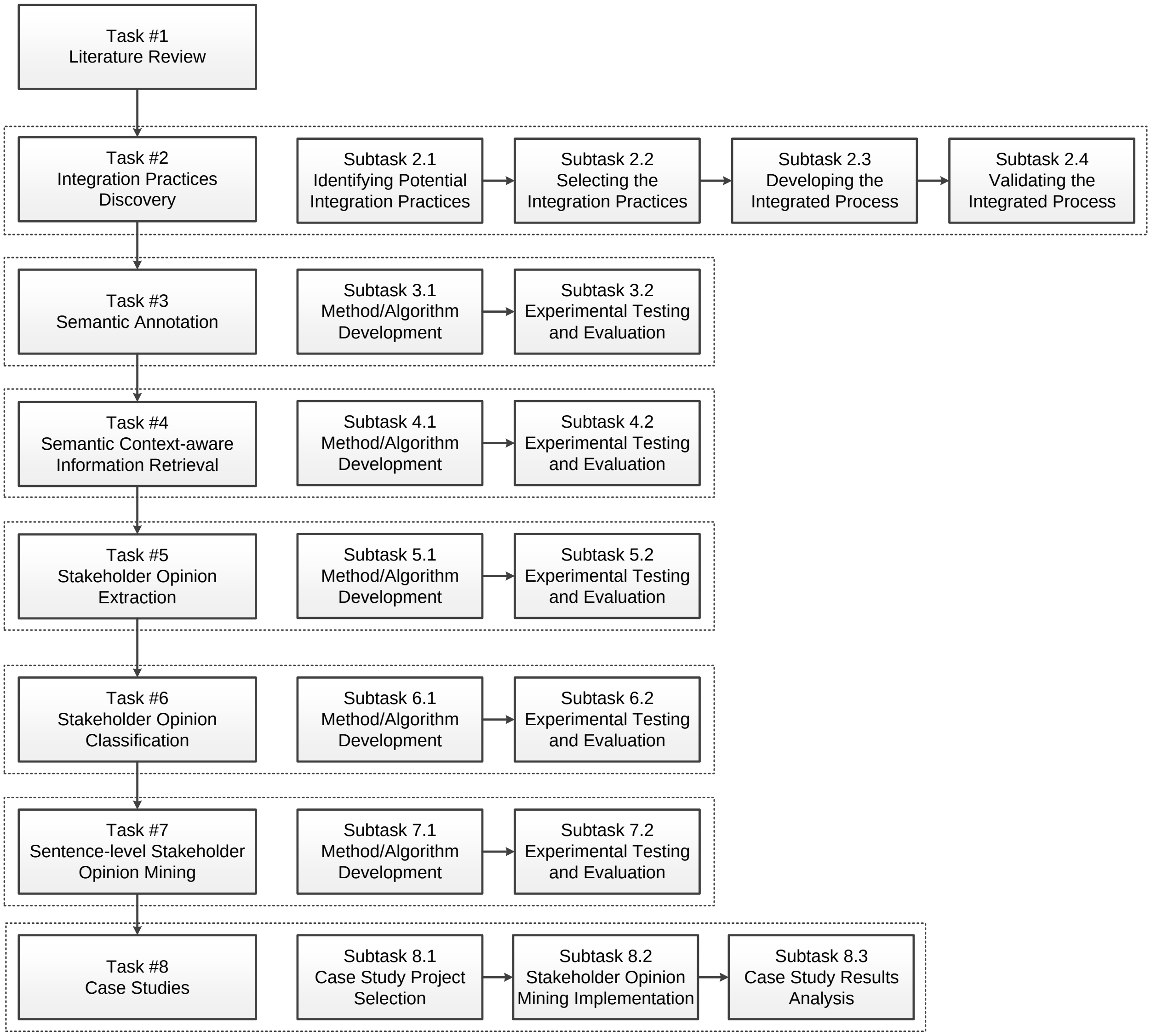

Task #1
Literature Review
Task #2
Integration Practices Discovery
Subtask 2.1
Identifying Potential Integration Practices
Subtask 2.2
Selecting the Integration Practices
Subtask 2.3
Developing the Integrated Process
Subtask 2.4
Validating the Integrated Process
Task #3
Semantic Annotation
Subtask 3.1
Method/Algorithm Development
Subtask 3.2
Experimental Testing and Evaluation
Task #4
Semantic Context-aware Information Retrieval
Subtask 4.1
Method/Algorithm Development
Subtask 4.2
Experimental Testing and Evaluation
Task #5
Stakeholder Opinion Extraction
Subtask 5.1
Method/Algorithm Development
Subtask 5.2
Experimental Testing and Evaluation
Task #6
Stakeholder Opinion Classification
Subtask 6.1
Method/Algorithm Development
Subtask 6.2
Experimental Testing and Evaluation
Task #7
Sentence-level Stakeholder Opinion Mining
Subtask 7.1
Method/Algorithm Development
Subtask 7.2
Experimental Testing and Evaluation
Task #8
Case Studies
Subtask 8.1
Case Study Project Selection
Subtask 8.2
Stakeholder Opinion Mining Implementation
Subtask 8.3
Case Study Results Analysis

Figure 1.1 – Research Methodology and Tasks

1.5.1 Research Task #1 – Literature Review

A comprehensive literature review was conducted in eleven primary domains: integrating NEPA and transportation planning processes, epistemology, semantic annotation, semantic similarity measures, information retrieval, document ranking models, stakeholder sentimental analysis and opinion mining, stakeholder opinion extraction, machine learning algorithms, text classification, and latent Dirichlet allocation. The following points provide a summary of the literature review in each of these domains.

- Integrating NEPA and transportation planning process: the literature review focused on (1) existing federal guidelines and efforts on integrating NEPA and transportation planning processes, (2) existing integration guidelines and efforts in other states (with focus on Florida, Colorado, Indiana, and Maine), and (3) existing transportation planning and NEPA processes in Illinois.
- Epistemology: the literature review focused on existing research on epistemology and its application in the construction domain.
- Semantic annotation: the literature review focused on existing research and methods for ontology-based semantic annotation including shallow semantic and deep semantic approaches.
- Semantic similarity measures: the literature review focused on existing research on semantic similarity measures that assess the semantic similarity between two concepts in a given semantic model.
- Information retrieval: the literature review focused on existing research and methods for context-aware information retrieval, including a review of relevant efforts in the construction and transportation domains.
- Document ranking models: the literature review focused on basic concepts of document ranking models and their applications including the vector space model and the statistical language model.
- Stakeholder sentiment analysis and opinion mining: the literature review focused on existing research on stakeholder opinion mining including document-level, sentence-level, and aspect-

level analysis, as well as lexicon-based, supervised machine learning-based, and unsupervised machine-learning-based approaches.

- Stakeholder opinion extraction: the literature review focused on existing methods and research on stakeholder opinion extraction including language rule-based, topic model-based, and supervised machine learning-based approaches.
- Machine learning: the literature review focused on the main types of machine learning algorithms and their characteristics and applications.
- Text classification: the literature review focused on existing methods for supervised machine learning-based text classification, with especial focus on multilabel text classification.
- Latent Dirichlet allocation: the literature review focused on the concept of latent Dirichlet allocation (LDA), and the collapsed Gibbs sampling method for inferencing distributions for LDA models.

1.5.2 Research Task #2 – Discovery of Integration Practices

This research task aimed to discover the integration practices for integrating the NEPA process into the state DOT and MPO planning process for large-scale highway projects in Illinois. This research task includes four primary subtasks.

1.5.2.1 Subtask 2.1 – Identifying Potential Integration Practices

A list of potential integration practices were identified based on two main sources: (1) a comprehensive literature review of existing processes, as well as existing integration guidelines and efforts, and (2) input from experts from relevant federal, state, and metropolitan planning and regulatory agencies. A comprehensive literature review of IDOT planning, MPO planning, and NEPA processes was conducted. Existing documents/studies that describe and/or evaluate the current practices of linking/integrating NEPA and transportation planning processes in other states were studied. Other relevant regulations and information resources including NEPA regulations, the Federal Highway Administration (FHWA) Planning and Environment Linkages (PEL) initiative and its related publications, and reports by the National Cooperative Highway Research Program (NCHRP) were also reviewed. Special emphasis was placed on reviewing integration efforts by states that have recently developed guidance on how to integrate transportation planning and NEPA processes, including Colorado, Florida, Indiana, and Maine. Expert inputs were gathered through unstructured meetings/interviews with eight experts from IDOT, FHWA, and MPOs. The purpose of those meetings was to gain a better understanding of the existing processes in Illinois and the appropriateness of potential integration practices.

1.5.2.2 Subtask 2.2 – Selecting the Integration Practices

This subtask focused on selecting the set of integration practices for the state of Illinois based on expert opinion from relevant federal, state, and metropolitan planning, regulatory, and resource agencies. A set of one-to-one expert interviews were conducted to collect data about current conditions and solicit expert opinion on the potential integration practices. Interviews were conducted face-to-face or online, with the preferred method being face-to-face and online only used if so desired by the respondent. Each interview consisted of two parts. The first part of the interview covered a presentation about the motivation and scope of the research. In the second part of the interview, respondents were asked to complete a questionnaire. Four main expert groups were identified – based on their responsibilities in the transportation planning and NEPA processes: (1) IDOT districts, (2) MPOs, (3) resource agencies, and (4) IDOT Central Office (Office of Planning and Programming and Bureau of Design and Environment) and FHWA. Accordingly, four questionnaires were designed.

1.5.2.3 Subtask 2.3 – Developing the Integrated IDOT-MPO-NEPA Planning Process

The results of the expert interviews were reviewed and discussed through unstructured meetings with eight experts from IDOT, FHWA, and MPOs. The purpose of those meetings was to (1) review the recommended practices in terms of their feasibility and applicability in Illinois, and (2) solicit recommendations on developing the implementation details of the recommended practices. Based on the results of the interviews and expert input, a final set of recommended integration practices were identified – considering feasibility and applicability – and were formulated into a coherent process workflow (and called Integrated IDOT-MPO-NEPA Planning Process). To represent the integrated process, a process flowchart was developed and each process was described in terms of process inputs, outputs, and actors. To facilitate the future evaluation of the integrated process, a set of performance measures were also identified based on literature review and recommendations from unstructured meetings with the eight experts.

1.5.2.4 Subtask 2.4 – Validating the Integrated IDOT-MPO-NEPA Planning Process

A second round of one-to-one, face-to-face interviews, which targeted the same group of experts in the first round, was conducted to validate the integrated process and evaluate its specific implementation details. To solicit expert feedback in an efficient manner, a draft guidance document describing the integrated process and a questionnaire was developed and sent to each of the interviewees two weeks prior to the interview date to allow interviewees sufficient time for review. Each interview consisted of two parts. The first part included a detailed presentation of the integrated process. In the second part of the interview, the interviewees were asked to complete the questionnaire to gather their opinions on the proposed integrated process. A six-point Likert scale was used to record the responses, with 6 being the most favorable (6=strongly agree, 5=agree, 4 = Somewhat Agree, 3 = Somewhat Disagree, 2 = Disagree, 1 = Strongly Disagree). For each question, the respondents were also asked to specify any recommendations or suggestions they may have on the specific implementation details of the integrated process. For all responses, mean, standard deviation, and median scores were calculated.

1.5.3 Research Task #3 – Development of Method and Algorithm for Semantic Annotation

This task aimed to develop a semantic annotation (SA) method and algorithm for automatically annotating textual documents with TPER-domain-specific concepts. This task focused on annotating webpages in the TPER domain with functional process context concepts, which describe the subprocesses of the TPER process. The functional process context is a subconcept of the document context, which is a subconcept of the epistemic context in the TPER epistemology. The TPER epistemology is a semantic model for representing and reasoning about information and information retrieval in the TPER domain. This research task was divided into two primary subtasks.

1.5.3.1 Subtask 3.1 – Method/Algorithm Development

This subtask focused on experimenting with different SA algorithms and semantic similarity measures to develop an SA method and algorithm for automatically annotating textual documents with TPER-domain-specific concepts. Two main types of SA algorithms were developed and comparatively evaluated: shallow SA and deep SA algorithms. The shallow SA algorithms mainly used syntactic features to annotate the text with concepts in the TPER epistemology. In developing the proposed shallow SA algorithm, three main algorithms were tested and evaluated: (a) using original concept terms (from the TPER epistemology) only; (b) conducting syntactic concept expansion on original concept terms; and (c) conducting both syntactic concept expansion and concept filtering and domain-specific concept expansion. The deep SA algorithms used the TPER epistemology for annotation and involved deep semantic analysis. In developing the proposed deep SA algorithm, eight different semantic similarity measures were tested and evaluated.

1.5.3.2 Subtask 3.2 – Experimental Testing and Evaluation

This subtask focused on testing and evaluating the developed methods and algorithms using well-established information retrieval metrics: mean precision and mean average precision. Mean precision for a set of concepts is the arithmetic mean of the precision values of the concepts. Precision, here, is defined as the ratio of the number of documents annotated correctly over the total number of documents annotated. The mean average precision is the mean of the average precision scores of each concept. For each concept, average precision is the average precision values at the ranks where correctly annotated documents occur (i.e., at the ranks where recall changes). These measures were calculated based on a comparison of the experimental results with a manually-developed gold standard. For a concept, each document in the collection was ranked based on the annotation weight, and the mean precision and mean average precision values at the top 10, 20, 30, 40, and 50 annotated documents were calculated.

1.5.4 Research Task #4 – Development of Method and Algorithm for Semantic Context-Aware Information Retrieval

This task aimed to develop a semantic context-aware information retrieval method and algorithm for retrieving relevant information in the TPER domain. This task focused on retrieving relevant information in the TPER domain based on the document context, which is a subconcept of the epistemic context in the epistemology. This research task was divided into two primary subtasks.

1.5.4.1 Subtask 4.1 – Method/Algorithm Development

This subtask focused on developing a semantic-based context-aware information retrieval method and algorithm and was composed of two main steps:

- Development of context-based relevance assessment method: A new context-based relevance assessment method was proposed to improve both context representation and context-based relevance estimation for enhancing the relevance of the retrieved results for decision making.
- Integration of the proposed relevance assessment method into document ranking models: The proposed context-based relevance assessment method was integrated into the vector space model (VSM) and the statistical language model (SLM), in order to (1) evaluate the effectiveness of the proposed relevance assessment method, (2) determine which method, the context-enhanced VSM or the context-enhanced SLM, results in a better information retrieval performance in the TPER domain.

1.5.4.2 Subtask 4.2 – Experimental Testing and Evaluation

This subtask focused on testing and evaluating the developed method and algorithm using well-established information retrieval metrics – precision, recall, and mean average precision. Precision, here, is defined as the ratio of the number of relevant documents retrieved over the total number of documents retrieved. Recall, here, is defined as the ratio of the number of relevant documents retrieved over the total number of relevant documents. The mean average precision is the mean of the average precision scores of each query. For each query, average precision is the average precision values at the ranks where relevant documents are retrieved (i.e., at the ranks where recall changes). These measures were calculated based on a comparison of the experimental results with a manually-developed gold standard. To develop the gold standard, a set of queries that represent the needs of transportation practitioners were developed by the experts in the TPER domain. For each query, the top 50 ranked documents retrieved by the proposed semantic ranking algorithms were pooled together and judged by domain experts. The relevant documents in the pool were considered relevant documents for the evaluation, and the rest of the documents in the pool together with the unjudged documents were considered irrelevant for the evaluation.

1.5.5 Research Task #5 – Development of Method and Algorithm for Stakeholder Opinion Extraction

The research task aimed to develop a stakeholder opinion extraction method and algorithm for automatically extracting subject, concern, and opinion expressions from stakeholder comments on large-scale highway projects to support aspect-level stakeholder opinion mining in the TPER domain. The research task was divided into two primary subtasks.

1.5.5.1 Subtask 5.1 – Method/Algorithm Development

This subtask focused on experimenting with different machine learning algorithms and performance improvement strategies to develop a stakeholder opinion extraction method and algorithm that could achieve sufficient performance. There are three main types of opinion extraction approaches: language rule-based approaches, topic model-based approaches, and supervised ML-based approaches. A language rule-based approach utilizes pre-defined rules to extract opinion-related expressions. A topic model-based approach generates opinion-related expressions through representing the comments with a mixture of topic models. A supervised ML-based approach learns to extract opinion-related expressions from manually-labeled data. In this thesis, the supervised ML-based approach was adopted because of its expected best performance and ability to extract fine-grained and precise information. Five different ML algorithms and three different types of features (syntactic, dependency, and semantic features) were comparatively evaluated, and a set of language rules were utilized to further improve the extraction performance.

1.5.5.2 Subtask 5.2 – Experimental Testing and Evaluation

This subtask focused on testing and evaluating the developed method and algorithm using well established information extraction metrics – precision, recall, and F1 measure. Precision, here, is defined as the ratio of the number of correctly extracted expressions (subject expressions, concern expressions, and opinion expressions) over the total number of extracted expressions. Recall, here, is defined as the ratio of the number of correctly extracted expressions over the total number of expressions that should be extracted. F1 measure is the harmonic mean of precision and recall. These measures were calculated based on a comparison of the experimental results with a manually-developed gold standard.

1.5.6 Research Task #6 – Development of Method and Algorithm for Stakeholder Opinion Classification

The research task aimed to develop an unsupervised ML-based method and algorithm for stakeholder opinion classification to support aspect-level opinion mining in the TPER domain. The research task focused on classifying the extracted opinion tuples into one or more concern categories and one sentiment category. This research task was divided into two primary subtasks.

1.5.6.1 Subtask 6.1 – Method/Algorithm Development

This subtask focused on developing an unsupervised ML-based method and algorithm for classifying the subject, concern, and opinion expressions (opinion tuples) that were extracted in Task 5. The developed method can automatically create labeled training through iteratively generating opinion tuple clusters, based on keywords, for each classification category. For clustering, semantic similarities between opinion tuples were captured through opinion semantic vectors, which were learned from a text corpus using a word-embedding model. The developed method then utilized a supervised ML classifier to learn from the automatically-created training data to classify the aspect-level opinion tuples into different concern categories (e.g., mobility and accessibility, air quality, transportation safety, etc.) and into one sentiment category (supportive, unsupportive, or neutral). In developing the proposed method, four different types of opinion semantic vectors, two supervised ML algorithms, and different cluster percentages used for training were comparatively evaluated.

1.5.6.2 Subtask 6.2 – Experimental Testing and Evaluation

This subtask focused on testing and evaluating the developed method and algorithm using well established text classification metrics – precision, recall, and F1 measure. Precision, here, is defined as the ratio of the number of correctly classified opinion tuples over the total number of classified opinion tuples. Recall, here, is defined as the ratio of the number of correctly classified opinion tuples over the total number of opinion tuples that should be classified. F1 measure is the harmonic mean of precision and recall. These measures were calculated based on a comparison of the experimental results with a manually-developed gold standard.

1.5.7 Research Task #7 – Development of Method and Algorithm for Sentence-level Stakeholder Opinion Mining

The research task aimed to develop a sentence-level stakeholder opinion mining method and algorithm for classifying sentences from stakeholder comments on large-scale highway projects into concern and sentiment categories. Compared to the tuple-based method (Research Tasks 5 and 6), the sentence-level method offers an alternative approach when a sentence-level analysis is sufficient. This research task was divided into two primary subtasks.

1.5.7.1 Subtask 7.1 – Method/Algorithm Development

This subtask focused on developing an unsupervised ML-based method and algorithm for classifying the sentences from stakeholder comments. The developed method can automatically create pseudo training data through latent Dirichlet allocation (LDA)-based concern labeling and lexicon-based sentiment labelling. The developed method then utilized a supervised ML classifier to learn from the automatically-created pseudo training data to classify the comment sentences into different concern categories (e.g., mobility and accessibility, air quality, transportation safety, etc.) and into one sentiment category (supportive, unsupportive, or neutral). In developing the proposed method, the effect of varying the size of the pseudo training data was comparatively evaluated.

1.5.7.2 Subtask 7.2 – Experimental Testing and Evaluation

This subtask focused on testing and evaluating the developed method and algorithm using well established text classification metrics – precision, recall, and F1 measure. Precision, here, is defined as the ratio of the number of correctly classified comment sentences over the total number of classified comment sentences. Recall, here, is defined as the ratio of the number of correctly classified comment sentences over the total number of comment sentences that should be classified. F1 measure is the harmonic mean of precision and recall. These measures were calculated based on a comparison of the experimental results with a manually-developed gold standard.

1.5.8 Research Task #8 – Case Studies of Stakeholder Opinion Mining

This task aimed to analyze stakeholder comments from a set of case study projects to gain a better understanding of stakeholder opinions, and how they could be similar or different across different stakeholder groups. This research task was divided into three subtasks.

1.5.8.1 Case Study Project Selection

This subtask aimed to select the case study projects. Three large-scale highway projects were selected due to their impacts on the surrounding environment, and the availability of their stakeholder comments. For each project, two primary stakeholder groups were identified: agency and government, and individual and public organization. For each project, stakeholder comments received during their respective planning processes were extracted from project reports (pdf format). These include comments provided through project websites, public hearings, emails, and social media. The extracted stakeholder comments were stored in a .txt format local file with textual content cleaned, and figures and tables removed.

1.5.8.2 Stakeholder Opinion Mining Implementation

This subtask aimed to implement the opinion mining method for each project, which involved (1) classifying the comment sentences into one or more concern categories, and into one sentiment category; and (2) aggregating sentence-level concern and sentiment labels to form the comment-level label set.

1.5.8.3 Case Study Results Analysis

This subtask aimed to answer the following research questions through analyzing the stakeholder opinions for the case study projects. For each project, (1) What are the support levels of the stakeholders to the project? (2) What are the concerns of the stakeholders? (3) What are the negative concerns of the stakeholders? (4) What are the similarities and differences – in support levels, concerns, and negative concerns – across the different stakeholder groups?

To answer the abovementioned research questions, the distributions of sentiments and concerns across the two stakeholder groups and three projects were analyzed. The two stakeholder groups were also compared, in terms of support levels, concerns, and negative concerns.

1.6 Contribution

1.6.1 Intellectual Merit

The thesis research contributes to the body of knowledge in six primary ways. First, it identifies the appropriate integration practices for Illinois through an in-depth investigation of existing planning processes and potential integration practices, develops a process model for the integrated IDOT-MPO-NEPA planning process, and provides well-defined guidance on the implementation and evaluation of the integrated process. Second, this research offers a domain-specific, deep semantic annotation algorithm for automatically annotating documents with concepts in the TPER epistemology. Third, this research offers a domain-specific, context-aware information retrieval algorithm for retrieving relevant documents in the TPER domain. Fourth, this research offers a domain specific, supervised ML-based information extraction method for automatically extracting subject, concern, and opinion expressions from stakeholder comments on large-scale highway projects. Fifth, this research offers a domain-specific, unsupervised ML-based stakeholder opinion classification method that supports early identification of stakeholder concerns and support levels. Sixth, this research offers a domain-specific, supervised ML-based stakeholder opinion mining method for classifying comment sentences on large-scale highway projects into one or more concern categories, and into one sentiment category. Seventh, this research offers a better understanding of stakeholder opinions and their similarities and differences among different stakeholder groups through analyzing stakeholder comments from three large-scale highway projects.

More detailed discussions of the intellectual merit and contribution to the body of knowledge are provided in Chapter 10.

1.6.2 Broader Impact

The research outcomes are expected to result in three primary broader impacts. First the implementation of the integrated IDOT-MPO-NEPA process could improve interagency coordination and communication, enable early identification of potential environmental issues and early consideration of avoidance/mitigation measures, and facilitate the use of early planning data/decisions in subsequent NEPA studies. All would result in improving the decision-making process, reducing duplication of work, and enhancing project delivery in terms of time and cost. Second, the implementation of the semantic, context-aware information retrieval methods could improve the ability of transportation practitioners to find the right information, at the right time, for the task at hand. This would help support TPER decision making and would reduce the time that agency employees spend to look for information in unstructured documents. Third, the implementation of the stakeholder opinion mining method could improve the ability of transportation practitioners to identify the concerns and support levels of the stakeholders early in the project development process. This would help avoid (or reduce) late identification of concerns and opposition, and accordingly would help reduce design changes and duplication of effort. All these potential outcomes are expected to reduce the time and cost of the TPER process, and further avoid delays in the project development process.

1.7 Publications

The thesis contains material published in the following papers and report:

- El-Gohary, N., Liu, L., El-Rayes, K., and Lv, X. (2014). *ICT Project R27-132 Incorporating NEPA into IDOT and MPO Planning Processes*, Illinois Center for Transportation, Rantoul, IL.
- Lv, X., and El-Gohary, N. (2015). "Semantic annotation for context-aware information retrieval for supporting the environmental review of transportation projects." *2015 ASCE International Conference on Computing in Civil Engineering (IWCCE)*, University of Texas at Austin, Austin, TX.
- Lv, X., and El-Gohary, N.M. (2016). "Text analytics for supporting stakeholder opinion mining for large-scale highway." *2016 International Conference on Sustainable Design, Engineering and Construction (ICSDEC)*, University of Arizona, Tempe, AZ.
- Lv, X., and El-Gohary, N.M. (2016). "Semantic-based information retrieval for supporting project decision making." *2016 International Conference on Computing in Civil and Building Engineering (ICCBE)*, Osaka University, Osaka, Japan.
- Lv, X., and El-Gohary, N.M. (2016). "Semantic annotation for supporting context-Aware information retrieval in the transportation project environmental review domain." *Journal of Computing in Civil Engineering*, 10.1061/(ASCE)CP.1943-5487.0000565, 04016033.
- Lv, X., and El-Gohary, N.M. (2016). "Discovering context-specific integration practices for integrating NEPA into statewide and metropolitan project planning processes." *Journal of Construction Engineering and Management*, 10.1061/(ASCE)CO.1943-7862.0001129 , 04016056.
- Lv, X., and El-Gohary, N.M. (2016). "Enhanced context-based document relevance assessment and ranking for improved information retrieval to support environmental decision making" *Journal of Advanced Engineering Informatics*, 30(4), 737 - 750.
- Lv, X., and El-Gohary, N.M. (2017). "Stakeholder opinion classification for supporting large-scale transportation project decision making." 2017 *International Workshop on Computing for Civil Engineering (IWCCE)*, University of Washington, Seattle, WA.

CHAPTER 2: LITERATURE REVIEW

2.1 Integrating NEPA and Transportation Planning Process

2.1.1 Federal Integration Efforts

According to the National Environmental Policy Act (NEPA), transportation projects are required to go through an environmental review process to evaluate their impact on the environment. Although the NEPA enhances the consideration of potential environmental consequences, and brings the general public and other stakeholders into the transportation decision-making process, it has received increasing criticism for "resulting in frequent delays in the development of important projects designed to improve the safety and operating conditions of a region's transportation system" (Larson et al. 2011). Since NEPA's enactment in 1969, the time it takes to complete an EIS has nearly tripled: in 1970s, a typical EIS took an average of 2.5 years to perform, and currently it takes an average of 6.5 years to complete (Barberio et al. 2008a). The reasons for EIS project delay, according to a series of research studies conducted by FHWA (FHWA 2000), include "a lack of funding or priority, stakeholder and/or local opposition, insufficient political support, project complexity, changes in agency priorities, environmental concerns expressed by resource agencies, and other issues inherent in the NEPA process itself" (Barberio et al. 2008a). In recognition of these reasons and potentials for improvement, the federal government has developed several guidance for integrating NEPA into transportation project planning processes.

Section 1309 of the Transportation Equity Act for the 21st Century (TEA-21) initiated the federal guidance for integrating the NEPA process into the state DOT and MPO planning processes in 1998 (USGPO 1998); it mandated the development and implementation of a coordinated environmental review process especially for projects that require the preparation of an environmental impact statement (EIS) pursuant to NEPA requirements

In 2007, Section 6002 of the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU) established a new environmental review process for transportation projects that require an EIS in order to promote efficient project management and enhanced interagency coordination (USGPO 2007).

The most up-to-date federal guidance for integrating the NEPA process into state DOT planning and MPO planning processes is provided in Section 1301-1323 of the Moving Ahead for Progress in the 21st Century Act (MAP-21) (USGPO 2013). These three sections introduce programmatic approaches to promote greater linkages between planning and environmental review process, and establish frameworks for setting deadlines for decision making during the environmental review process considering conflict resolution and penalties for agencies that fail to make a decision (USGPO 2013).

2.1.2 States Integration Efforts

In response to the federal guidance, a number of states have conducted extensive research studies on how to integrate the NEPA process into their transportation project planning processes, and developed detailed and formal guidelines on how to implement and evaluate the integrated process. The Colorado Department of Transportation (CDOT) conducted the Strategic Transportation, Environmental, and Planning Process for Urban Places (STEP-UP), where a Geographic Information System (GIS)-based tool to identify and assess the environmental impacts and a methodology to conduct regional cumulative effects assessment were developed (FHWA 2007a; MacDonald and Lidov 2007). The Florida Department of Transportation (FDOT) developed the

Efficient Transportation Decision-Making (ETDM) process, which utilized the Environmental Screening Tool (EST), an internet-accessible interactive database for documenting project changes, evaluating impacts, and communicating project details to agencies and the public (FDOT 2006; FHWA 2007b). The Indiana Department of Transportation (INDOT) developed a streamlined procedure for planning and environmental analysis to eliminate the duplication of activities between planning studies and subsequent environmental analysis carried out under NEPA (FHWA 2007c; INDOT and FHWA 2007). The Maine Department of Transportation (MaineDOT) developed the Maine's Integrated Transportation Decision Making (ITD) process for integrating existing project review processes to eliminate duplication of efforts, and the process is designed for projects that require an EIS or EA (FHWA 2002; FHWA 2007d).

2.1.3 Transportation Planning and NEPA Processes in Illinois

In Illinois, a transportation project originates from a project concept that aims to solve specific regional or statewide transportation needs (IDOT 2010). At the regional level, through the MPO planning process, the 16 MPOs in Illinois develop the MPO's Long-Range Transportation Plan (LRTP) as the guidance for decision making and identify important projects that satisfy the regional transportation needs for inclusion in the plan (IDOT 2006). At the statewide level, through the IDOT planning process, IDOT districts receive project proposals from MPOs and other regional planning agencies and select priority projects for inclusion in the Multi-Year Program (MYP), where the funding of each project is specified (IDOT 2007). Once the project is funded, a project group is assigned to supervise the project development. If the project involves more than one alternative corridor within a regional area, a corridor/feasibility study may be conducted to investigate all feasible corridors (IDOT 2010). To determine the specific alignments, profiles, and major design features of the proposed project, Phase I (design) studies are conducted. If the project has (or may have) a significant environmental impact, then a NEPA study is conducted as part of Phase I studies. The NEPA study focuses on the environmental considerations of the project – including impacts on the social, cultural, and economic resources, as well as natural resources – and follows one of two types of processes for large-scale transportation projects (CEQ 2007; CEQ 2006): (1) EIS process, if the project is identified to have significant environmental impact, and (2) Environmental Assessment (EA) process, if there are uncertainties about whether the project will have significant environmental impact.

2.2 Epistemology

Epistemology is a branch of philosophy about the nature of knowledge (Muis 2004); it aims to investigate how the knowledge of a particular domain is created and disseminated (Steup 2011). Recently, the need for epistemological understanding and modeling to support effective knowledge management has been recognized; "it is necessary to understand the broad epistemological spectrum that can enable effective utilization of computerized systems for knowledge management" (Jayatilaka and Lee 2003). Specifically, in the areas of information systems and information retrieval, researchers have highlighted the epistemological nature of information systems and information retrieval processes. An information system acts as an "epistemology, not just extending human abilities but offering a new approach to knowing" (Broman 2014). "Classic models of search indicate that the information retrieval process involves: the identification of a need; the search to meet that need; the evaluation of results towards the need. This process has parallels in models of 'epistemic beliefs' " (Knight 2013). Modeling the IR process as a knowing process would enable better representation of the information retrieval process in terms of its epistemic context [context of knowing (i.e., searching), context of knower (i.e., user), and context of knowledge (i.e., document)]. Such representation could better facilitate context-aware information retrieval.

In the construction domain, the need for epistemological understanding and modeling in supporting construction informatics research has also been recently emphasized (El-Diraby 2012). Most recently, an epistemology-based semantic model for facilitating domain-specific, context-aware retrieval of information about sustainable construction practices was proposed (Zhang and El-Gohary 2015).

2.3 Semantic Annotation

As the corner stone of context-aware information retrieval, semantic annotation (SA) is the process of assigning the semantic descriptions to the entities in the text (Kiryakov et al. 2004); it can bridge the gap between the computer-understandable knowledge and the extensive human natural-language materials (Li and Bontcheva 2007). Current researchers have focused on three different types of SA (Castells et al. 2007; Fernandez et al. 2011): (1) Statistical approaches, which identify groups of words that commonly appear together, based on a statistical model, and use these word groups as semantic descriptions. For example, using modified latent semantic analysis (LSA), Ozcan and Aslandogan (2005) identified the concepts in a domain-specific corpus for query expansion, and achieved significant improvement in the precision of information retrieval; (2) Linguistic conceptualization approaches, which take advantage of linguistic resources like WordNet or thesauri to enhance document indexing. For example, Boubekeur et al. (2010) proposed a concept-based document indexing approach using WordNet; they assigned concepts extracted from WordNet to document words based on the overlapping degree between a WordNet synset and the local context, and measured the concept weight based on semantic relatedness and concept frequency; and (3) Ontology-based approaches, which link the concepts in the ontological model with the text, and provide a much more detailed and densely populated concept space in the form of an ontology (Fernandez et al. 2011).

In comparison to ontology-based approaches, statistical and linguistic conceptualization approaches (1) are commonly based on shallow and sparse conceptualizations, (2) usually consider very few types of relations between concepts, and (3) usually allow for low information specificity levels (Castells et al. 2007).

2.4 Semantic Similarity Measures

Semantic similarity (SS) measures determine how much two concepts are similar according to a given semantic model, and are "becoming intensively used for most applications of intelligent knowledge-based and semantic information retrieval systems" (Slimani 2013). In this work, SS measures are used to identify the match between concepts in a user's query and concepts in a document. A number of measures have been proposed to assess the SS between pairs of concepts based on an ontology (or concept hierarchy). The measures can be classified into the following three categories: path-based measures, node-based measures, and combined measures.

Path-based measures estimate the SS between two concepts based on the shortest path between the two concepts (Zhang et al. 2007). Some popular measures are: (1) Wu and Palmer (1994) SS measure, which utilizes the shortest path between the two concepts and their most informative subsumer (MIS) in the hierarchy (MIS is the lowest concept that can be a parent for both the two concepts). This measure assumes that the shortest path length and the depth of MIS are equally important in assessing SS, which may not be suitable for hierarchically specific concepts; (2) Leacock and Chodorow (1998) measure, which transforms the shortest path distance into a similarity measure and normalizes it by the maximum depth of the hierarchy. The drawback of this measure is that it assumes the links in the hierarchy represent uniform semantic distances, which is typically not true because a semantic distance is also affected by concept specificity and the density of the subhierarchy; (3) Li et al. (2003) SS measure, which combines the length of the shortest path between the two concepts and the depth of their MIS in the hierarchy through two parametric functions. This SS measure uses two parameters to set the importance of concept specificity and shortest path, which allows for tuning/optimizing the contributions of these two features based on empirical results; and (4) Mao and Chu (2007) SS measure, which utilizes the shortest path between the two concepts and their descendant concepts. This SS measure is built on the assumption that a concept is less similar to its grandparent than to its parent in the hierarchy, and takes the generality of concepts into account by considering the number of their descendants. This SS measure works effectively when evaluating a concept and its descendant concepts, but not for evaluating sibling concepts.

Node-based measures estimate the SS between two concepts based on the information content (IC) of the concept nodes. Some popular measures are: (1) Resnik (1995) SS measure, which is based on the intuition that the SS between two concepts is the extent to which they share common information and utilizes the IC of two concepts' MIS to measure the shared information. This measure has limited capability to differentiate concept pairs that have the same MIS, because it only considers the MIS of two concepts; (2) Jiang and Conrath (1997) SS measure, which combines the IC of each concept in addition to the IC of their MIS. It improves Resnik (1995) SS measure by introducing the IC of two concepts to differentiate concept pairs that have the same MIS; and (3) Lin (1998) SS measure, which utilizes the ratio between the IC of two concepts' MIS and the sum IC of the two concepts. It reflects not only how much common information the two concepts have, but also how much different information they have.

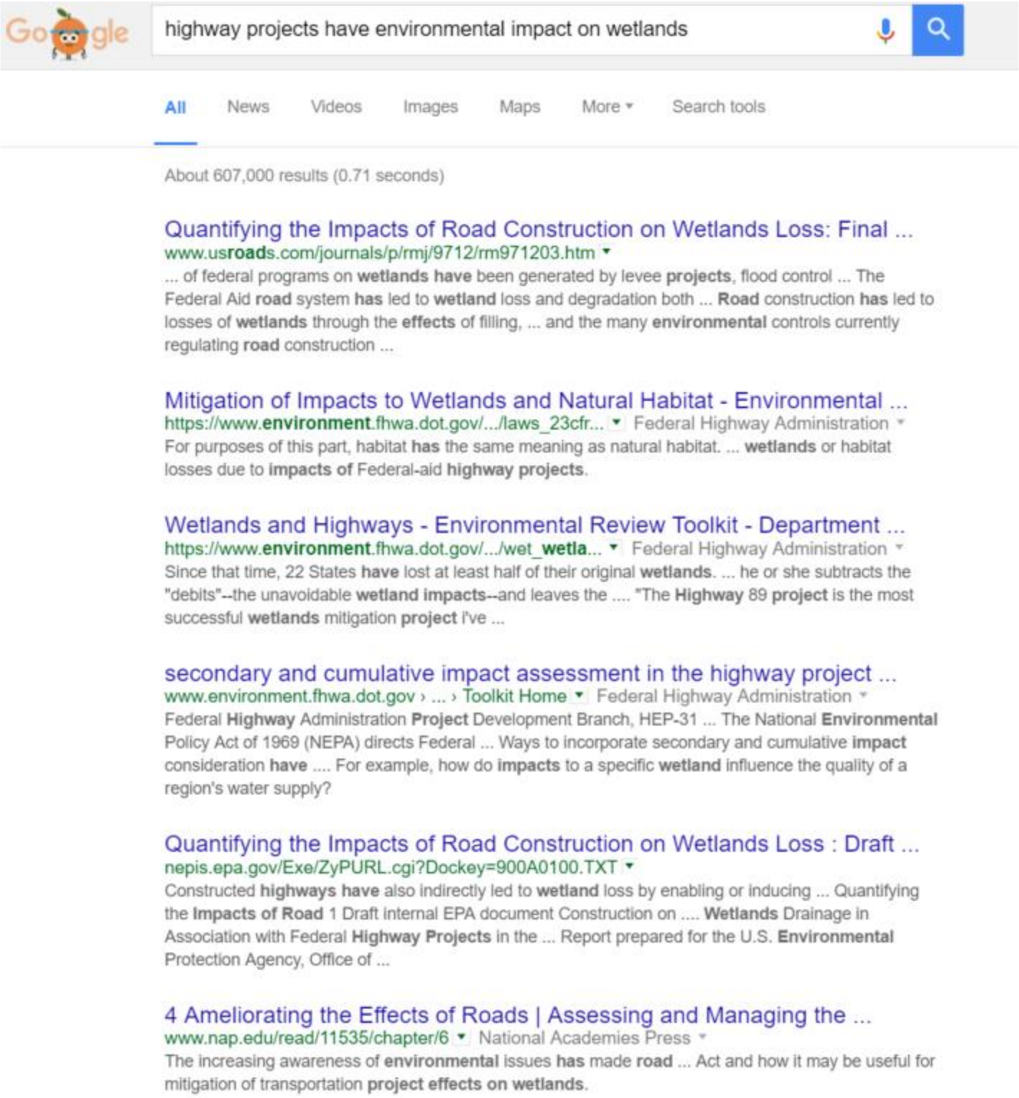
Al-Mubaid and Nguyen (2006) SS measure is a combined measure that utilizes the shortest path between two concepts and the common specificity (CSpec) of the two concepts. The CSpec of two concepts is the difference between the maximum IC of all concepts in the hierarchy and the IC of the two concepts' MIS. CSpec indicates how much common information two concepts share, and the lower their CSpec is the more information they share. Compared with other path-based and node-based measures, Al-Mubaid and Nguyen (2006) SS measure has the following advantages: (1) It is based on not only the distance between two concepts in the hierarchy (path feature), but also the amount of common information they share (node feature); and (2) It uses two parameters to set the importance degrees of the path feature and the node feature, which allows for tuning/optimization of parameters based on empirical results.

Different types of SS measure have their own advantages and limitations, and previous studies (Stevenson and Greenwood 2005; Petrakis et al. 2006; Budanitsky and Hirst 2006; Meng et al. 2013) indicate that no single type can outperform the other types in all applications. For example, Petrakis et al. (2006) found that the Leacock and Chodorow (1998) measure achieved the best performance in concept term stemming compared with other existing measures; while Stevenson and Greenwood (2005) found the Jiang and Conrath (1998) SS measure to be the best measure for conducting pattern induction for information extraction. Because the performances of SS measures vary from application to application, and most of the applications are based on WordNet or ontologies in the bio-medical domain, it is necessary to evaluate the performances of the different types of SS measures in SA in the TPER domain.

2.5 Information Retrieval

Information retrieval (IR) is the process of finding material (usually documents) of an unstructured nature (usually text) that satisfies the user's information need within large collections (Manning et al. 2009). The current IR systems mostly build on keyword-based content representation and query processing techniques, which provide limited capabilities for incorporating content semantics and contextual information into the retrieval process (Fernandez et al. 2011). Due to this limitation, the current IR systems can be very ineffective when handling context-sensitive tasks, such as search that involves terms of multiple meanings. To overcome this limitation of keyword-based IR systems, context-aware IR – which aims to integrate search technologies and knowledge about the query and the context into a single framework in order to provide the most relevant answer for a user's information need – has been recognized as a long-term challenging goal in the IR research domain (Allan et al. 2003; Ozcan and Aslandogan 2005; Kara et al. 2012; Chauhan et al. 2013).

For the transportation environmental review domain, as indicated by recent studies (ICT 2014), the ineffectiveness of current IR systems are aggravated when searching for relevant information to support decision making for the domain. For example, the following use case scenario provides an illustrative example: an environmental specialist (user's role) from IDOT is working on a new toll road corridor (project type) that affects nearby wetlands (affected resource) in northeastern Illinois (project location), he/she would like to find similar projects that also affect wetlands and how their environmental impacts are evaluated, and he/she searchers Google for "highway projects have environmental impact on wetlands". Figure 2.1 shows the first result page that was retrieved by Google. All the retrieved results in the first page only provide general information about evaluating environmental impacts on wetlands, such as guidance on quantifying the impacts on wetland loss (first and fifth results), and mitigation measures for the impacts on wetlands (second result); and none of them provide the specific project examples that the environmental specialist needs to retrieve. To improve the retrieval results, he/she enhances the query and searches Google for "Illinois tollway projects have environmental impact on wetlands". Figure 2.2 shows the first result page using the enhanced query. Although the third and sixth retrieved results provide information on the specific projects the environmental specialist are looking for, other results only provide general information such as guidance on wetland restoration (first and second results) and the environmental studies manual (fifth). The highly context-sensitive nature of the transportation environmental review process and of the searching process of related information makes it difficult to retrieve satisfactory results using conventional IR systems. The searching process of the environmental review relevant information is sensitive to the context of the domain knowledge (e.g., project type, project location, environmental review type, affected resources, etc.), the context of the user (e.g., user role, user task at hand, user profile), and the context of the searching process (e.g., searching location, searching environment, searching device). For example, in the above use case scenario, the information on the desired highway projects is sensitive to the project type, project location, environmental resources affected, and user role. An enhanced semantic-based document ranking method is needed to help retrieve more relevant results by adapting to these various contexts.



highway projects have environmental impact on wetlands

About 607,000 results (0.71 seconds)

Quantifying the Impacts of Road Construction on Wetlands Loss: Final ...
www.us**road**s.com/journals/p/rmj/9712/rm971203.htm
... of federal programs on **wetlands have** been generated by levee **projects**, flood control ... The Federal Aid **road** system **has** led to **wetland** loss and degradation both ... **Road** construction **has** led to losses of **wetlands** through the **effects** of filling, ... and the many **environmental** controls currently regulating **road** construction ...

Mitigation of Impacts to Wetlands and Natural Habitat - Environmental ...
https://www.**environment**.fhwa.dot.gov/.../laws_23cfr... Federal Highway Administration
For purposes of this part, habitat **has** the same meaning as natural habitat. ... **wetlands** or habitat losses due to **impacts of** Federal-aid **highway projects**.

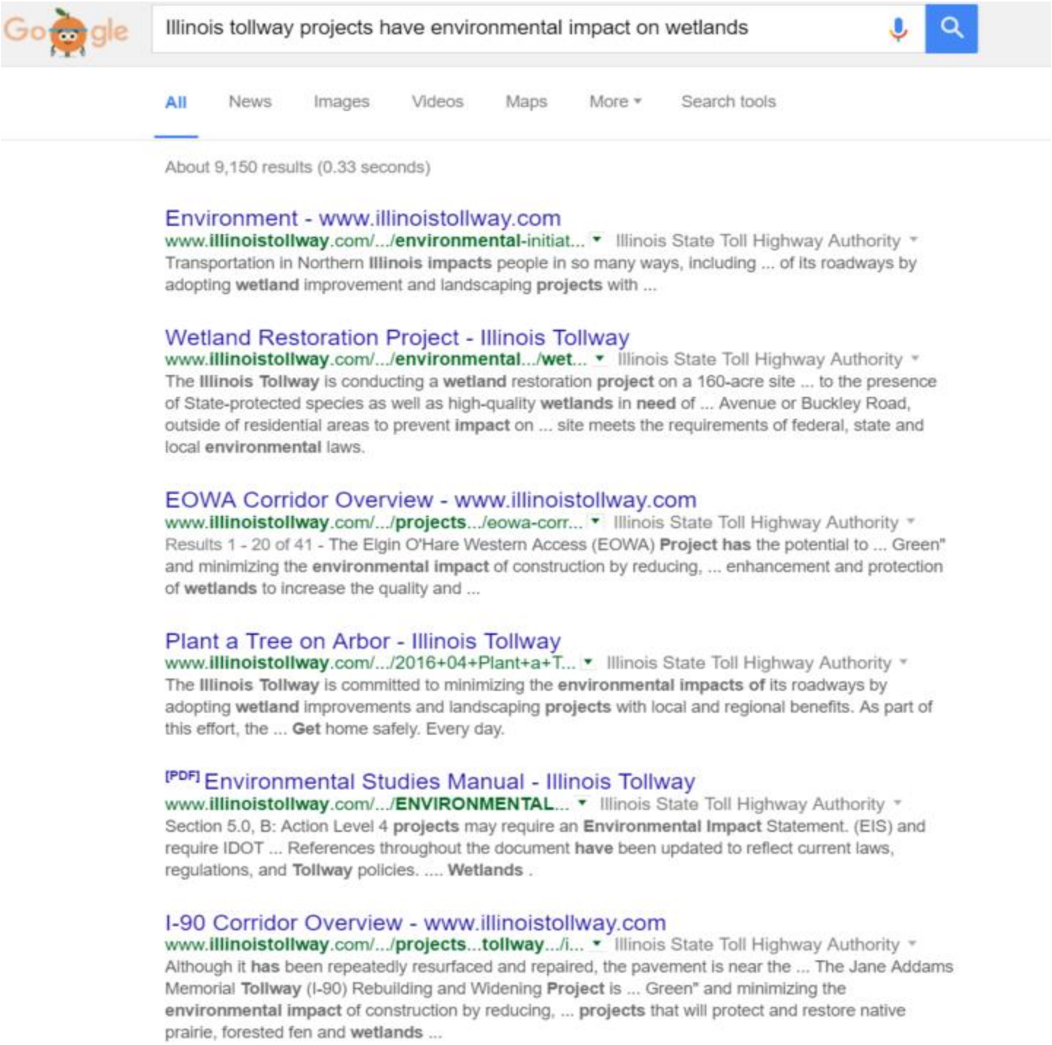
Wetlands and Highways - Environmental Review Toolkit - Department ...
https://www.**environment**.fhwa.dot.gov/.../wet_**wetla**... Federal Highway Administration
Since that time, 22 States **have** lost at least half of their original **wetlands**. ... he or she subtracts the "debits"--the unavoidable **wetland impacts**--and leaves the .... "The **Highway** 89 **project** is the most successful **wetlands** mitigation **project** i've ...

secondary and cumulative impact assessment in the highway project ...
www.environment.fhwa.dot.gov › ... › Toolkit Home Federal Highway Administration
Federal **Highway** Administration **Project** Development Branch, HEP-31 ... The National **Environmental** Policy Act of 1969 (NEPA) directs Federal ... Ways to incorporate secondary and cumulative **impact** consideration **have** .... For example, how do **impacts** to a specific **wetland** influence the quality of a region's water supply?

Quantifying the Impacts of Road Construction on Wetlands Loss : Draft ...
nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=900A0100.TXT
Constructed **highways have** also indirectly led to **wetland** loss by enabling or inducing ... Quantifying the **Impacts of Road** 1 Draft internal EPA document Construction on .... **Wetlands** Drainage in Association with Federal **Highway Projects** in the ... Report prepared for the U.S. **Environmental** Protection Agency, Office of ...

4 Ameliorating the Effects of Roads | Assessing and Managing the ...
www.nap.edu/read/11535/chapter/6 National Academies Press
The increasing awareness of **environmental** issues **has** made **road** ... Act and how it may be useful for mitigation of transportation **project effects on wetlands**.

Figure 2.1 – The First Result Page Retrieved by Google Using the Example Original Query



Illinois tollway projects have environmental impact on wetlands

About 9,150 results (0.33 seconds)

Environment - www.illinoistollway.com
www.illinoistollway.com/.../environmental-initiat... Illinois State Toll Highway Authority
Transportation in Northern Illinois impacts people in so many ways, including ... of its roadways by adopting wetland improvement and landscaping projects with ...

Wetland Restoration Project - Illinois Tollway
www.illinoistollway.com/.../environmental.../wet... Illinois State Toll Highway Authority
The Illinois Tollway is conducting a wetland restoration project on a 160-acre site ... to the presence of State-protected species as well as high-quality wetlands in need of ... Avenue or Buckley Road, outside of residential areas to prevent impact on ... site meets the requirements of federal, state and local environmental laws.

EOWA Corridor Overview - www.illinoistollway.com
www.illinoistollway.com/.../projects.../eowa-corr... Illinois State Toll Highway Authority
Results 1 - 20 of 41 - The Elgin O'Hare Western Access (EOWA) Project has the potential to ... Green" and minimizing the environmental impact of construction by reducing, ... enhancement and protection of wetlands to increase the quality and ...

Plant a Tree on Arbor - Illinois Tollway
www.illinoistollway.com/.../2016+04+Plant+a+T... Illinois State Toll Highway Authority
The Illinois Tollway is committed to minimizing the environmental impacts of its roadways by adopting wetland improvements and landscaping projects with local and regional benefits. As part of this effort, the ... Get home safely. Every day.

[PDF] Environmental Studies Manual - Illinois Tollway
www.illinoistollway.com/.../ENVIRONMENTAL... Illinois State Toll Highway Authority
Section 5.0, B: Action Level 4 projects may require an Environmental Impact Statement. (EIS) and require IDOT ... References throughout the document have been updated to reflect current laws, regulations, and Tollway policies. .... Wetlands .

I-90 Corridor Overview - www.illinoistollway.com
www.illinoistollway.com/.../projects...tollway.../i... Illinois State Toll Highway Authority
Although it has been repeatedly resurfaced and repaired, the pavement is near the ... The Jane Addams Memorial Tollway (I-90) Rebuilding and Widening Project is ... Green" and minimizing the environmental impact of construction by reducing, ... projects that will protect and restore native prairie, forested fen and wetlands ...

Figure 2.2 – The First Result Page Retrieved by Google Using the Example Enhanced Query

2.6 Document Ranking Models

A document ranking model provides the basic notion of what it means for a document to be relevant to a query. Among the many different document ranking models proposed in the literature, the vector space model (VSM) and the statistical language model (SLM) are the most studied and widely used. The VSM is a similarity-based model that assumes that the relevance of a document to a query is correlated with the similarity between the query and the document at some level of representation (Aggarwal and Zhai 2012). In the VSM, a document and a query are represented as two vectors of terms, which are typical words and phrases. Each term is assigned a weight that reflects its “importance” to the document or the query. This model measures the relevance of a document to a query as the similarity between the query vector and document vector. The cosine similarity and the inner-product between the two vectors are often used as the similarity measures (Ceri et al. 2013).

The SLM is a probabilistic model that assumes that the documents in a collection should be ranked by the decreasing probabilities of their relevance to a query (Singhal 2001). A document is generally viewed as a sample from a language model, which estimates the distribution of words in a given language. Based on this assumption, this model measures the relevance of a document to a query as the likelihood that the query was generated based on the estimated language model of each document (Zhai 2008).

2.7 Stakeholder Sentiment Analysis and Opinion Mining

Sentiment analysis (also called opinion mining) is the task of detecting, extracting, and classifying opinions, sentiments, and attitudes concerning different topics from unstructured stakeholder opinions (Montoyo et al. 2012; Ravi and Ravi 2015). A stakeholder opinion is a piece of text that expresses the attitude(s) of a stakeholder towards a target object, such as a product or service (e.g., a highway project in the context of this work). The target object under evaluation is defined as an entity, and an entity could have several aspects representing its features (Liu and Zhang 2012). Based on the level of analysis granularity, sentiment analysis can be conducted at three different levels: document level, sentence level, and aspect level (Liu 2012). Document-level sentiment analysis aims to determine whether the whole opinion document expresses positive, negative, or neutral sentiment based on the assumption that each document expresses opinion(s) about a single entity (Pang et al. 2002; Medhat et al. 2014). Sentence-level sentiment analysis, on the other hand, analyzes each sentence in an opinion document, and identifies the sentiment of each sentence. At the aspect level, sentiment analysis aims to identify the aspects (i.e., features of a target object) that are covered in a comment sentence, and discover the commenter's sentiment attitude(s) towards each of these aspects. Compared with the document and sentence levels, aspect-level sentiment analysis performs a finer-grained analysis, which could help discover the specific issues stakeholders like or dislike.

Sentence-level sentiment analysis is based on the simple assumption that a sentence expresses a single sentiment from a single commenter (Liu 2012). Three major approaches have been proposed for sentence-level sentiment analysis: lexicon-based approach, supervised machine learning (ML)-based approach, and unsupervised ML-based approach. The lexicon-based approach determines the sentiment orientation of a sentence by summing up the sentiment scores of all opinion words in the sentence using a pre-defined opinion lexicon (Liu and Zhang 2012). The supervised ML-based approach relies on supervised ML algorithms to solve the sentence-level sentiment analysis as a text classification problem and requires the representation of sentences using syntactic, linguistic, and/or semantic features (Medhat et al. 2014). Some of the commonly used algorithms are naive Bayers, maximum entropy, support vector machines, and neural networks. The unsupervised ML-based approach learns to classify sentences from unlabeled training data using an unsupervised ML algorithms such as topic modelling and text clustering algorithms (Pang and Lee 2008).

Aspect-level sentiment analysis consists of two main tasks: aspect extraction and aspect sentiment classification. Aspect extraction is the process of extracting entity (subject), aspect, and/or opinion expressions from the stakeholder opinions. Aspect sentiment classification is the process of determining whether the opinions on the different aspects are positive, negative, or neutral. There are three main aspect-level sentiment classification approaches that have been proposed in recent years: lexicon-based, supervised machine learning (ML)-based, and unsupervised ML-based approaches.

Lexicon-based approaches utilize an opinion lexicon, which consists of opinion words and/or phrases, and a set of rules to determine the sentiment orientation of aspects. For example, Hu and Liu (2004) built an opinion lexicon by propagating seed words with known semantic orientation through searching the WordNet synonym/antonym graph. An aspect is assigned with the sentiment (positive or negative) of the majority of sentiment-bearing adjectives in the sentence, or the sentiment of the closest sentiment-bearing adjective when the number of positive and negative adjectives is the same. Ding et al. (2008) extended Hu and Liu (2004)'s opinion lexicon with a context-dependent opinion-word list and opinion-idiom rules to mark opinion words and phrases; and used a set of linguistic rules to handle but-clauses and opinion shifters such as negation words like not, never, and none. The opinion orientation expressed on each aspect is represented as an opinion score and computed using an opinion aggregation function. One main limitation of the lexicon-based approach is that the performance of sentiment classification depends largely on the quality of the opinion lexicon, and opinion lexicons usually do not cover all types of expressions that convey or imply opinions.

Supervised ML-based approaches treat aspect-level sentiment analysis as a text classification problem, and utilize supervised ML algorithms to classify stakeholder opinions through learning from labeled training data. For example, Choi and Cardie (2008) adapted the simple bag-of-words approach for sentiment classification of opinion tuples by incorporating structural inference motivated by compositional semantics into the learning procedure. An SVM-based classifier was used to determine the polarity of an expression in a two-step process, where the polarity of the constituents are determined first, and then combined recursively to form the polarity of the whole expression based on inference rules. Yu et al. (2011) conducted sentiment classification on product reviews through applying an aspect ranking-based weighting scheme to an SVM-based classifier. The aspect ranking algorithm considers both the aspect frequency and the contribution of commenters' opinion on specific aspects to the overall opinion, and gives higher weights on important aspects and sentiment terms that modify these aspects. Akhtar et al. (2017) used a particle swarm optimization (PSO)-based method for feature selection and ensemble learning to conduct aspect-level sentiment analysis. The ensemble classifier combines the outputs of three base classifiers – maximum entropy (ME), conditional random fields (CRF), and SVM – using a majority voting.

Unsupervised ML-based approaches rely on unsupervised ML algorithms such as topic modelling to learn the sentiment orientation of aspects from unlabeled training data. Propeseu and Etzioni (2007) identified potential opinion phrases from searching the vicinity of each explicit aspect, where the vicinity is measured using syntactic dependencies. The semantic orientation of each explicit aspect is then determined using a relaxation labeling technique, which finds the most likely polarity labels for extracted sentiment phrases while satisfying many types of local constraints, such as conjunctions and disjunctions. Mei et al. (2007) proposed a probabilistic model to capture aspects and sentiments simultaneously in Weblogs. The proposed topic-sentiment mixture model assumes a blog article is generated by sampling words from a mixture model of a background language model, a set of topic (aspect) language models, and two (positive and negative) sentiment language models. Poria et al. (2016) proposed the Sentic latent Dirichlet allocation (LDA) framework to better capture semantics in aspect-level sentiment analysis through integrating semantic similarity into the calculation of word distributions in the LDA algorithm.

2.8 Stakeholder Opinion Extraction

Stakeholder opinion extraction – an important element of stakeholder aspect-level sentiment analysis (or opinion mining) – is the process of extracting entity, aspect, and/or opinion expressions from unstructured stakeholder opinions. There are three main stakeholder opinion extraction approaches that have been proposed in recent years: language rule-based approach, topic model-based approach, and supervised machine learning (ML)-based approach.

The language rule-based approach extracts opinion-related expressions using predefined rules, which capture the contextual patterns and/or grammatical relations between the terms in the text (Zhang and Liu 2014). For example, Hu and Liu (2004) proposed an extraction method based on association rules, which finds frequent aspects through frequent nouns and noun phrases, and identifies infrequent aspects using dependency relations between aspects and opinion words. Qiu et al. (2011) developed the double-propagation method to extract aspects and opinions simultaneously based on direct dependency relations. Poria et al. (2014) exploited common-sense knowledge and sentence-dependency trees to detect both explicit and implicit aspects from product reviews. One limitation of the rule-based approach is the adaptability of language rules, because the performance of rules depends largely on the document collection; rules that work well on one collection may not work well on another.

The topic model-based approach assumes that the stakeholder opinions are generated through mixtures of topic models, and each topic model is a unigram language model that represents a type of aspect. For example, Mukherjee and Liu (2012) developed two joint aspect-opinion models for extracting and categorizing aspects at the same time given user-provided seed words. Chen et al. (2014) proposed an aspect extraction framework to extract more coherent aspects by exploiting the knowledge automatically learned from online reviews. One major limitation of the topic model-based approach is that it can only find some general aspects, and has difficulty in finding fine-grained or precise aspects.

The supervised ML-based approach learns to extract aspects from manually labeled data. Some methods utilized sequence models, which treat aspect extraction as a sequence-labeling task. For example, Jin et al. (2009) utilized a lexicalized hidden Markov model (HMM), which incorporates linguistic features such as part-of-speech and lexical patterns to extract aspects from product reviews. Jakob and Gurevych (2010) evaluated the performance of a conditional random fields (CRF)-based method for aspect extraction in a single and cross-domain environment. Shariaty and Moghaddam (2011) employed CRF for identifying product aspects and proposed a technique for defining and filtering features to enhance the performance. Toh and Wang (2014) developed an aspect-based sentiment analysis system, which extracts aspect terms from product reviews using CRF and a combination of general features (e.g., part-of-speech tags) and open features (e.g., WordNet Taxonomy). Shu et al. (2017) proposed a lifelong learning method to improve the aspect extraction performance by enabling CRF to leverage the knowledge gained from previous extraction results from other domains. Less commonly, other researchers used supervised learning models that treat aspect extraction as a binary or multi-class classification task. For example, Ghani et al. (2006) used both supervised and semisupervised algorithms to extract attribute and value pairs from product descriptions. Yu et al. (2011) trained a one-class SVM algorithm to identify aspects in the candidate noun phrases extracted from pros-and-cons consumer reviews. Poria et al. (2016) proposed a deep learning approach to tag words in opinion sentences as either aspect or non-aspect based on a 7-layer deep convolutional neural network (CNN).

Compared with the language rule-based approach and the topic model-based approach, the supervised ML-based approach typically has the best extraction performance, and is able to extract fine-grained and precise information..

2.9 Machine Learning Algorithms

Eight ML algorithms are reviewed in this section, because they were used in this research. Five ML algorithms were used in developing the proposed stakeholder opinion extraction method: HMM, maximum entropy markov model (MEMM), CRF, structured perceptron (SP), and SVM-HMM. HMM is a probabilistic model for sequential data, which models the joint distribution of both the observation and the labels. It assumes that the current label only depends on its previous label, and the current observation only depends on the current label (Zhang and Liu 2014). HMM has two limitations. First, it does not allow the use of overlapping features that are not independent of each other, such as part-of-speech and dependency features. Second, it maximizes the likelihood of the observation and label sequences, while the sequence labeling task is to maximize the likelihood of the label sequence given the observation sequence (McCallum et al. 2000).

MEMM overcomes the above-mentioned limitations through directly modelling the probability of the label sequence given the observation sequence. MEMM assumes that the probability of transitioning to a particular label depends on the current observation and the previous label, thus allowing the use of multiple, non-independent features of observations (McCallum et al. 2000). However, because MEMM uses a per-state exponential model, it may suffer from the label bias problem. For example, the term “toll” can be the beginning of a subject expression such as “toll road”, or the beginning of a concern expression such as “toll fees”. If the term “toll” is more likely to be the beginning of a subject expression in the training data, the MEMM algorithm would always label “toll” as “S-B” (beginning of a subject expression) regardless of the following terms.

CRF is a type of discriminative undirected probabilistic graphical model, which defines a conditional probability distribution over label sequences given a particular observation sequence, rather than a joint distribution over both label and observation sequences (Sutton and McCallum 2012). The conditional nature of the CRF results in the relaxation of the independence assumptions required by HMMs, which allows the use of arbitrary, non-independent features. Additionally, CRF avoids the label bias problem of MEMM by having a single exponential model for the joint probability of the entire label sequence (Lafferty et al. 2001).

SP is an extension of the conventional perceptron to handle structured prediction problems (Collins 2002), and has many desirable properties. First, it does not require the calculation of a partition function, which is necessary for other structured prediction algorithms (Lafferty et al. 2001). Second, it is also robust to approximate inference, which is often required for problems where the search space is too large and where strong structural independence assumptions are insufficient (Collins 2002).

SVM-HMM is a maximum margin model that aims to maximize the difference between the correct label sequence and its closest incorrect label sequence (Tsochantaridis et al. 2004). As a combination of HMM and SVM, it inherits the advantages of both algorithms: modeling the label sequence and observation in a discriminative approach which can account for overlapping features, and allowing the use of kernel functions to learn nonlinear discriminant functions (Altun et al. 2003)..

Three ML algorithms were used in developing the proposed stakeholder opinion classification method: SVM, backpropagation for multilabel learning (BP-MLL), and convolutional neutral networks (CNN). SVM is a classification algorithm which aims to maximize the margin between the hyperplanes defined by the different classes of data (Basu et al. 2003). As a maximum margin model, SVM has good ability in handling high dimensional features and sparse document vectors. It is, therefore, widely used for text classification (Joachims, 2002). SVM also allows the use of kernel function to transform a feature space to a higher dimensional space in order to solve non-linear separation problems (Joachims, 2002).

BP-MLL (Zhang and Zhou 2006) is a multilabel classification algorithm that builds on the neural network model. It formulates multilabel classification problem as a neural network with multiple output nodes for each label and extends the backpropagation algorithm through designing a new error function that is able to capture the characteristics of multilabel learning.

A CNN is a type of deep, feed-forward artificial neural networks that consists of convolution layers, pooling layers, and fully-connected layers (Liu and Zhang 2018). The CNN model is a useful algorithm for text classification because the convolutional and pooing layers allow the model to find local indicators (e.g., sequence of words) of class memberships regardless of their position in the text.

2.10 Text Classification

Text classification (TC) is reviewed in this section, because the opinion classification problem was be formulated as a TC problem. In general, a TC problem aims to classify documents (like stakeholder comments in this research) into one or more categories (Aggarwal and Zhai 2012). A TC problem could be categorized as a multilabel or single-label classification problem (Tsoumakas and Katakis 2007). Multilabel TC can assign more than one label to a document, while single-label TC can only assign one label to each document. Depending on the number of unique labels to be assigned, a single-label TC problem can be further categorized as a binary classification problem or multiclass classification problem. Existing multilabel classification methods can be grouped into two main categories: problem transformation methods (PTMs), and algorithm adaptation methods (AAMs).

PTMs assume the labels are independent with each other, and transform the multilabel TC problem into multiple single-label TC subproblems. If the number of labels to be assigned is *n*, then after the transformation, there would be *n* single-label classification subproblems (thus *n* classifiers) and *n* number of data sets (one data set for each label L_k). The single-label subproblems after transformation are commonly addressed using a binary classification approach. For each subproblem with label L_k, the binary classification approach treats the label L_k as the positive category and combines all other labels in a negative category. During the training process, each classifier is trained on the dataset to predict the corresponding label. During the testing process, each opinion tuple is judged by those *n* classifiers to decide whether to assign its corresponding label or not. All the assigned concern labels (by the *n* classifiers) form the final label set of this opinion tuple.

However, PTMs can create some disadvantages (Tsoumakas and Katakis 2007) as they ignore the label correlation by falsely assuming the labels are independent, and would fail when predicating certain combinations of labels, thus leading to undermined performance. PTMs can also show data imbalance problems when solving single-label subproblems using a binary classification approach. As a result of combining labels, the negative examples often outnumber the positive ones by a large margin, which affects the classification performance. To overcome these disadvantages, AAMs can cope with multilabel TC problems directly by modifying or extending available algorithms. For example, based on the traditional k-nearest neighbors (KNN) algorithm, the multilabel KNN (ML-KNN) algorithm (Zhang and Zhou 2007) first identifies the k-nearest neighbors in the training set for each unseen instance. After that, the algorithm utilizes the maximum a posteriori principle to determine the label set for the unseen instance based on statistical information gained from the label sets of these neighboring instances, such as the number of neighboring instances belonging to each possible category.

2.11 Latent Dirichlet Allocation

Topic models are statistical models for discovering topics that occur in a collection of documents (Blei 2012). The Latent Dirichlet Allocation (LDA) model (Blei et al. 2003) is a popular topic model that generates documents based on probabilistic rules. LDA is an unsupervised learning algorithm that describes a set of documents as a probabilistic mixture of distinct topics, where each topic is a probability distribution over the words in the document collection (Blei et al. 2003). Two types of methods can be used to learn these probability distributions: the variation inference methods, which approximate posterior distributions through expectation and maximization (EM)-based optimization (Hoffman et al. 2011), and Markov chain Monte Carlo (MCMC) methods, which approximate posterior distributions through random sampling using probabilistic rules (Darling 2011). The Gibbs sampling method is the most commonly used MCMC method for LDA. It includes two steps: initialization step and iteration step. At the initialization stage, each word in every document is randomly assigned to one of the K topics. After the initialization step, the iteration step is conducted to update the topic assigned to each word in the document based on the learned topic assignment distribution. At each iteration, the topic assignment distribution is learned assuming all topic assignments except for the current word are correct (Griffiths and Steyvers 2004).

CHAPTER 3: DISCOVERY OF INTEGRATION PRACTICES FOR INTEGRATING NEPA, STATE DOT, AND MPO PLANNING PROCESSES

3.1 Identifying Potential Integration Practices

A comprehensive literature review of IDOT planning, MPO planning, and NEPA processes was conducted. Existing documents/studies that describe and/or evaluate the current practices of linking/integrating NEPA and transportation planning processes in other states were also studied. Other relevant regulations and information resources including NEPA regulations, the FHWA's Planning and Environment Linkages (PEL) initiative and its related publications, and reports by the NCHRP were also reviewed. Special emphasis was placed on reviewing integration efforts by states that have recently developed guidance on how to integrate transportation planning and NEPA processes, including Colorado, Florida, Indiana, and Maine. Expert input was gathered through unstructured meetings/interviews with eight experts from IDOT, FHWA, and MPOs. The purpose of those meetings was to gain a better understanding of the existing processes in Illinois and the appropriateness of potential integration practices.

Based on the literature review and expert input, a list of 16 key integration practices were identified. The practices were classified into two main types: process-oriented integration practices and collaboration-oriented integration practices. Process-oriented integration practices are practices for integrating NEPA and transportation planning processes to allow for early and continuous agency participation; early identification of environmental, socioeconomic, and cultural impacts and concerns; reduced duplication of work; and reduced durations and efforts of project delivery. The following is a summarized description of the five main types of process-oriented integration practices:

- Practices related to the preparation of an MPO's LRTP: These practices are intended to achieve early coordination and engagement of resource agencies and IDOT during the preparation of the MPO's LRTP. Early participation of the resource agencies allows for early identification of critical environmental issues and avoidance of issues that could become fatal flaws at later stages of project development (FHWA 2007a; MacDonald and Lidov 2005).
- Practices related to environmental screening of projects during the MPO's planning process (planning screen): These practices are intended to enhance the effectiveness of planning by incorporating the consideration of environmental resources during the preparation of the MPO's LRTP (FDOT 2006). This is accomplished through an environmental screening of priority projects, which aims to estimate the environmental, socioeconomic, and cultural impacts of each project through a comparison of the location of the project and the locations of resources (FDOT 2006; FHWA 2007b).
- Practices related to the preparation of IDOT's MYP: These practices are intended to achieve early coordination and engagement of resource agencies while preparing the IDOT's MYP. The involvement of resource agencies at this stage could provide feedback on environmental issues for projects that are not included in the MPO's LRTP, and could help IDOT quickly identify participating agencies for subsequent NEPA studies (FDOT 2006; MacDonald and Lidov 2005).
- Practices related to environmental screening of projects during the IDOT's planning process (programming screen): These practices are intended to accelerate the subsequent NEPA process through evaluating potential environmental, socioeconomic, and cultural impacts and identifying project-specific environmental studies and analyses that are needed to satisfy NEPA (FDOT 2006). This is accomplished through a more comprehensive environmental

screening for a larger number of projects (including projects that were not evaluated in the planning screen) with detailed and updated project information (FDOT 2006; FHWA 2007b).

- Practices related to the preparation of a corridor/feasibility study: These practices are intended to reduce the duplication of work between corridor/feasibility studies and subsequent NEPA studies (FHWA 2011; INDOT and FHWA 2007). This is mainly achieved through conducting corridor/feasibility studies in compliance with NEPA requirements, including documentation requirements (FHWA 2011).

Collaboration-oriented integration practices aim to support the process-oriented practices by facilitating early, continuous, and in-depth interagency coordination and communication in order to support the integration of NEPA and transportation planning processes. Six collaboration-oriented integration practices were identified (FDOT 2006; FHWA 2007d; INDOT and FHWA 2007; MacDonald and Lidov 2007): (1) data management system, (2) memorandums of understanding (MOUs) and programmatic agreements (PAs), (3) interagency advisory group, (4) training and outreach, (5) designated coordinators at MPOs and IDOT districts, and (6) dedicated staff at resource agencies. A description of each practice is included in the Section 3.3.

3.2 Selecting Integration Practices

In order to select the appropriate integration practices for the state of Illinois, an expert survey was conducted. The purpose of the survey was to solicit (1) specific information on current conditions related to environmental analysis during the current transportation planning process (e.g., access to environmental screening tools), (2) the agreement level of experts on the potential effectiveness of the identified integration practices (e.g., conducting environmental screening of projects during the planning phase) in terms of enhancing efficiency of project delivery and improving interagency coordination, and (3) specific expert recommendations on how to implement the identified integration practices (e.g., the most suitable tool to use in conducting the screening). Although some identified potential integration practices have been proven to be effective in other states, whether they would be effective in Illinois depends on the Illinois context, in terms of current conditions, availability of resources, and willingness of planning and resource agencies to adopt new practices. For example, since planning agencies in Florida all have access to a GIS-based environmental screening tool and have reached an agreement with FDOT, FHWA, and resource agencies about their roles and responsibilities, the integration practice of conducting environmental screening during their planning process can be successfully implemented (FDOT 2005).

3.2.1 Questionnaire Design

A separate questionnaire was designed for each expert group because the responsibilities, and thus degree of expertise, of experts vary across each group. For example, IDOT districts have higher expertise in developing corridor studies than MPOs. Four main expert groups were, thus, identified – based on their responsibilities in the transportation planning and NEPA processes: (1) IDOT districts, (2) MPOs, (3) resource agencies, and (4) IDOT Central Office (Office of Planning and Programming and Bureau of Design and Environment) and FHWA. Accordingly, four questionnaires were designed.

Each questionnaire was composed of three main sections: respondent information, current conditions, and potential integration practices. Section 1 aimed to collect the following respondent information: name, contact information, agency he/she represents, and years of experience.

Section 2 aimed to solicit specific information on the current conditions related to environmental analysis during the current transportation planning process, including (1) what environmental screening tools MPOs and IDOT districts have access to; (2) what planning studies MPOs conduct, whether environmental considerations are taken into account when conducting these studies, and the reasons for not taking environmental considerations into account if that is the case; and (3) if MPOs and IDOT districts environmentally screen projects during their planning phases (i.e., during the preparation of the MPO's LRTP and IDOT's MYP, respectively), and if yes, at what point in the planning phase, for which types of projects (system maintenance, bridge maintenance, congestion mitigation, or system expansion projects), how frequent (for every one of those types of project, sometimes, or occasionally), and using which tool [Detailed Impact Review Tool (DIRT), Arch-GIS, or other].

Section 3, the main section, aimed at soliciting expert opinion about the potential effectiveness of the identified integration practices if implemented in the state of Illinois and recommendations on their implementation. All potential integration practices were listed and respondents were requested to rate their level of agreement with each practice on a six-point Likert scale, with 6 being the most favorable (6=strongly agree, 5=agree, 4=somewhat agree, 3=somewhat disagree, 2=disagree, 1=strongly disagree). For the process-oriented integration practices, for each potential practice, respondents were asked whether they agree that the practice could help reduce both the time and cost of the project development process. For the collaboration-oriented integration practices, for each practice, respondents were asked whether they agree that the practice is potentially effective in achieving early and continuous involvement and coordination. To solicit expert recommendations on the implementation, for practices about environmental screening, respondents were further asked about the recommended time to conduct the screening, the recommended tool to use for screening, and the recommended way(s) to disseminate the results of the screening. Respondents were also asked whether they recommend establishing and using standardized environmental criteria and metrics for conducting the screening. An open-ended question was also included at the end of Section 3 to ask respondents if they would like to recommend any other practices (other than the practices listed in the questionnaire) that could be potentially effective when implemented in Illinois.

3.2.2 Verifying the Questionnaire Design

To evaluate the effectiveness of the questionnaires, a pilot study was conducted with eight experts from IDOT, FHWA, and MPOs. The experts were requested to review each survey questionnaire, then to provide feedback on their format and content. Feedback was solicited on the various aspects of the questionnaires, such as question wording, response options and scale, clarity of the descriptions of the integration practices, and instructions to respondents. The questionnaires were then revised according to the feedback. For example, (1) for Likert scale questions asking about experts' opinion on potential integration practices, a "have no opinion" option was added in case a respondent was uncertain about the potential effectiveness of a practice, and (2) for questions that solicited information about the type(s) of projects currently being environmentally screened, definitions and examples of project types were added.

3.2.3 Survey Implementation

The expert survey was conducted from May to August 2013, in a one-on-one interview format. The interviews were conducted face-to-face or online. The preferred method was face-to-face; and online was only used if so desired by the respondent. Each interview consisted of two parts. The first part of the interview covered a presentation about the motivation and scope of the research. In the second part of the interview, respondents were asked to complete the questionnaire. The survey targeted experts who are involved in conducting, supervising, and/or coordinating planning and/or environmental studies (e.g., planning director, environmental study supervisor) at the following agencies: IDOT districts, MPOs, resource agencies, IDOT central office, and FHWA. These experts were targeted as they are more familiar with the roles and responsibilities of different agencies in the transportation planning process and/or NEPA process and can provide better feedback on the potential effectiveness of the identified integration practices. A total of 31 one-to-one survey interviews with experts from 29 agencies were conducted, including 21 face-to-face meetings and 10 online meetings. The respondent information of all the interviewees is summarized in Table 3.1.

Table 3.1 – Summary of Respondent Information for Integration Practices Evaluation Survey

Expert group	Number of respondents	Years of experience
IDOT Central Office and FHWA	4	All over 10 years
IDOT district	9	All over 10 years, except 1
MPO	12	All over 10 years
Resource agency	6	All over 10 years

3.2.4 Survey Results

For Likert-scale questions, the mean, standard deviation, median, and mode scores were calculated for each expert group and for all groups, and the results were interpreted based on the median scores.

In terms of the environmental analysis during the current transportation planning process in Illinois, nearly all respondent agencies (19 out of 21 districts and MPOs) have access to an environmental screening tool, with the majority of them having access to Arc-GIS (15 out of 19). For environmental screening, only a small number of the interviewed MPOs (3 out of 12) conduct an environmental screening during their planning process. All three MPOs conduct this screening once priority projects are selected for inclusion but prior to their inclusion in the LRTP, conduct it only for system expansion projects and only occasionally, and conduct it without using an environmental screening tool. For IDOT districts, the majority of the interviewed districts (6 out of the 9) conduct an environmental screening during their planning process. The majority of those six districts (5 out of the 6) conduct this screening once priority projects are selected for inclusion but prior to their inclusion in the MYP, with only one district screening candidate projects prior to their prioritization and selection. Half of the six districts screen only system expansion projects and only occasionally, while the other half screen all types of projects and every one of those types of projects. The tool used in screening varies across districts, where two use only DIRT, three use both DIRT and Arch-GIS, and one uses Project Monitoring Application (PMA).

For the process-oriented integration practices, the results are summarized in Table 3.2. The results indicate that respondents collectively "agree" or "somewhat agree" that the potential process-oriented practices could help reduce both the time and cost of the project development process. In terms of expert recommendations on the implementation of process-oriented practices, based on the median of responses from all expert groups, for Practice P3, experts recommended screening priority projects once they have been included in the MPO's LRTP. For Practice P6, they recommended screening a candidate project, at the district level, prior to the prioritization and selection of projects for inclusion in the MYP. For both Practice P3 and Practice P6, experts recommended (1) the use of a GIS-based tool, like ArcGIS, for screening, (2) establishing and using standardized environmental criteria and metrics for conducting the screening, and (3) disseminating the results of the screening by uploading and storing the results in a common database and by informing Phase I consultants, IDOT in-house staff, and resource agencies involved in the NEPA process of the results.

For the collaboration-oriented practices, the results are summarized in Table 3.3. The results indicate that respondents collectively "agree" or "strongly agree" that the proposed collaboration-oriented integration practices are potentially effective means for achieving early and continuous involvement and coordination.

Table 3.2 – Summary of Survey Results for Process-Oriented Integration Practices

Process-oriented integration practice	Mean	Standard deviation	Median	Mode	Overall opinion of respondents (based on median)
Practices related to the preparation of an MPO's LRTP					
P1: Ensuring early coordination between IDOT districts and MPOs while preparing the LRTPs by MPOs.	5.48	0.51	5	5	Agree
P2: Engaging resource agencies and soliciting their feedback on potential environmental issues during the preparation of the LRTPs by MPOs.	4.23	1.11	4	3	Somewhat agree
Practices related to the planning screen					
P3: Conducting environmental screening of projects during the planning phase (during the preparation of the MPO's LRTP).	3.58	1.00	4	3	Somewhat agree
P4: Establishing and using standardized environmental criteria and metrics for environmental screening during the planning phase.	4.80	0.37	5	5	Agree
Practices related to the preparation of IDOT's MYP					
P5: Engaging resource agencies and soliciting their feedback on potential environmental issues during the preparation of the MYP.	4.00	0.64	4	4,5	Somewhat agree
Practices related to the programming screen					
P6: Conducting environmental screening of projects during the programming phase (during the preparation of IDOT's MYP).	5.00	1.00	5	6	Agree
P7: Establishing and using standardized environmental criteria and metrics for environmental screening during the programming phase.	4.94	0.77	5	5	Agree
Practices related to the preparation of a corridor/feasibility study					
P8: Requiring corridor studies and feasibility studies to be conducted in compliance with NEPA requirements.	4.65	1.17	5	5	Agree
P9: Providing Phase I consultants involved in preparing corridor studies and/or feasibility studies with environmental screening information.	4.83	0.72	5	5	Agree

Table 3.3 – Summary of Survey Results for Collaboration-Oriented Integration Practices

Collaboration-oriented integration practice	Mean	Standard deviation	Median	Mode	Overall opinion of respondents (based on median)
P10: Establishing and using one common database for collecting, storing, updating, and accessing project data and environmental data.	5.00	0.45	5	5	Agree
P11: Developing memorandums of understanding (MOUs) and/or programmatic agreements (PAs) among agencies for supporting early and continuous involvement and coordination.	4.87	0.63	5	5	Agree
P12: Establishing interagency work groups, advisory groups, and/or committees for supporting early and continuous involvement and coordination.	4.55	0.78	5	5	Agree
P13: Providing agencies with a common understanding of one another's roles and responsibilities (e.g., through webinars).	5.10	0.54	5	5	Agree
P14: Designating a coordinator at every IDOT district to be responsible for the implementation of the streamlined NEPA/planning process and for interagency coordination.	4.68	0.65	5	5	Agree
P15: Designating a coordinator at every MPO to be responsible for the implementation of the streamlined NEPA/planning processes and for interagency coordination.	4.55	0.83	5	5	Agree
P16: Providing dedicated staff at resource agencies for cooperating and coordinating with IDOT/IDOT districts and MPOs.	5.29	0.94	6	6	Strongly agree

3.2.5 Validating the Reliability of the Survey Results

A Cronbach's alpha test was conducted on the Likert scale questions to validate the internal consistency (i.e., reliability) of the survey. Internal consistency indicates the extent to which all questions in a survey measure the same construct (i.e., opinion about integration practices, in this case). The Cronbach's alpha test is used to confirm the reliability of survey results and to ensure that the same results can be reasonably expected if a similar survey is conducted under similar circumstances (Buthelezi and Mkhize 2014). Alpha values of 0.7 or greater indicate adequacy/acceptability of internal consistency (Laerd Statistics 2013). The overall Cronbach's alpha value for the survey is 0.72, which indicates an adequate level of reliability.

3.3 Developing the Integrated IDOT-MPO-NEPA Planning Process

The results of the expert survey were further reviewed and discussed through unstructured meetings with eight experts from IDOT, FHWA, and MPOs. All potential practices were included in the initial set of recommended practices and were discussed during those meetings, because they all received positive expert feedback (i.e., an overall median of "somewhat agree", "agree", or "strongly agree"). The purpose of those meetings was to (1) review the recommended practices in terms of their feasibility and applicability in Illinois, and (2) solicit recommendations on developing the implementation details of the recommended practices. The final set of recommended practices were then identified considering their feasibility and applicability. All the initial recommended practices were included in the final set, except that the timing of Practice P6 was changed so that projects are screened once they are included in the IDOT's MYP. Subsequently, the final set of recommended integration practices were formulated into a coherent process workflow (and called Integrated IDOT-MPO-NEPA Planning Process). In order to develop the workflow, the final set of recommended process-oriented integration practices were integrated into the existing transportation planning processes, with the collaboration-oriented practices being ongoing efforts to foster early and continuous involvement and coordination across agencies and work groups. To facilitate the future evaluation of the integrated process, a set of performance measures were also identified.

3.3.1 Process-Oriented Integration Practices

Figure 3.1 shows a flowchart of the Integrated IDOT-MPO-NEPA Planning Process that summarizes the proposed subprocesses and their interactions, where an added or changed subprocess (i.e., a subprocess was added or elements of a subprocess were changed, in comparison to existing transportation planning processes) is highlighted with green color. Table 3.4 shows the inputs, outputs, responsible agencies, and other actors of a sample of the subprocesses. The following is a brief summary of the recommended process-oriented integrating practices:

- MPO’s LRTP Preparation: during the preparation of the LRTP, MPOs should coordinate with corresponding IDOT districts and solicit the feedback of resource agencies on potential environmental issues.
- Planning Screen: once priority projects are included in the MPO’s LRTP, the MPO, in cooperation with resource agencies, should conduct an environmental screen of those projects using a GIS-based tool and standardized environmental criteria and metrics. Once the screen is completed, the MPO should upload the data and the results of the screen in a common database and should inform the consultants, IDOT in-house staff, and resource agencies involved in the subsequent NEPA process of these data and results.
- IDOT’s MYP Preparation: during the preparation of the MYP, IDOT districts should solicit the feedback of resource agencies on potential environmental issues.
- Programming Screen: once large-scale highway projects are included in the IDOT’s MYP, the IDOT district, in cooperation with resource agencies, should conduct an environmental screen using a GIS-based tool and standardized environmental criteria and metrics. Once the screen is completed, the district should upload the data and the results of the screen in a common

database and should inform the consultants, IDOT in-house staff, and resource agencies involved in the subsequent NEPA process of these data and results.

- Corridor/Feasibility Studies Preparation: A corridor/feasibility study should be conducted in compliance with NEPA requirements, and prior to the start of the study consultants should be provided with the data and results of the planning and programming screens.

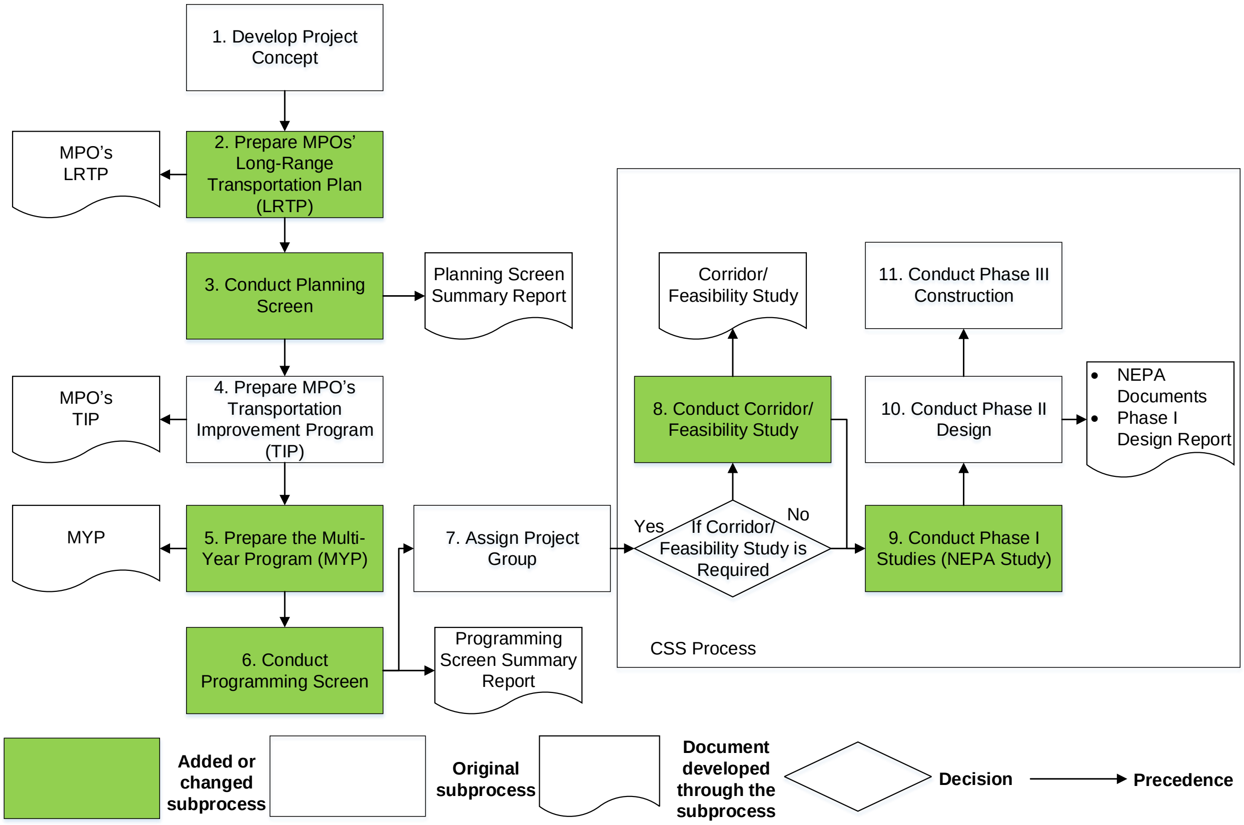

1. Develop Project Concept
MPO's LRTP
2. Prepare MPOs' Long-Range Transportation Plan (LRTP)
3. Conduct Planning Screen
Planning Screen Summary Report
MPO's TIP
4. Prepare MPO's Transportation Improvement Program (TIP)
MYP
5. Prepare the Multi-Year Program (MYP)
7. Assign Project Group
6. Conduct Programming Screen
Programming Screen Summary Report
Corridor/ Feasibility Study
11. Conduct Phase III Construction
8. Conduct Corridor/ Feasibility Study
10. Conduct Phase II Design
NEPA Documents
Phase I Design Report
Yes
If Corridor/ Feasibility Study is Required
No
9. Conduct Phase I Studies (NEPA Study)
CSS Process
Added or changed subprocess
Original subprocess
Document developed through the subprocess
Decision
Precedence

Figure 3.1 – Proposed Integrated IDOT-MPO-NEPA Planning Subprocesses

Table 3.4 – Inputs, Outputs, Responsible Agencies and Other Actors of Each Subprocess of the Integrated IDOT-MPO-NEPA Process (partial)

Subprocess	Inputs	Outputs	Responsible agencies	Other actors
Conduct planning screen	• Project, environmental, socioeconomic, and cultural data • Standardized criteria and metrics • Agency feedback	• Planning screen summary report	• MPO • IDOT district • Resource agencies	• Designated coordinator from MPO • Designated coordinator from IDOT district • Environmental coordinators • Interagency advisory group
Conduct programming screen	• Project, environmental, socioeconomic, and cultural data • Standardized criteria and metrics • Agency feedback	• Programming screen summary report	• IDOT district • MPO • Resource agencies	• Designated coordinator from MPO • Designated coordinator from IDOT district • Environmental coordinators • Interagency advisory group
Conduct corridor/ feasibility study	• Purpose and need • Planning screen summary report • Programming screen summary report • Project, environmental, socioeconomic, and cultural data • Agency feedback • Public feedback	• Corridor/ feasibility study report	• Project group • IDOT district • IDOT Central Office • MPO • Resource agencies • Consultants • General public	• Designated coordinator from MPO • Designated coordinator from IDOT district • Environmental coordinators • Interagency advisory group
Conduct Phase I studies (NEPA study)	• Purpose and need • Planning screen summary report • Programming screen summary report • Corridor/feasibility study report • Project, environmental, socioeconomic, and cultural data • Agency feedback • Public feedback	• Phase I studies plans and reports • NEPA documents	• Project group • IDOT district • IDOT Central Office • FHWA • Resource agencies • Consultants • General public	• Designated coordinator from MPO • Designated coordinator from IDOT district • Environmental coordinators • Interagency advisory group

3.3.2 Collaboration-Oriented Integration Practices

The following is a brief summary of the recommended collaboration-oriented integration practices:

- Common Database: establishing and using one common database for collecting, storing, updating, and accessing project data and environmental data, where data/feedback is provided and accessed by IDOT/IDOT districts, MPOs, resource agencies, and consultants.
- Designated Coordinators: designating a coordinator at every district and at every MPO to be responsible for the implementation of the Integrated IDOT-MPO-NEPA Process and for interagency coordination.
- Dedicated Staff at Resource Agencies: providing dedicated staff at resource agencies for cooperating and coordinating with IDOT (or IDOT districts) and MPOs.
- Interagency Advisory Groups: establishing interagency work groups, advisory groups, and/or committees for supporting early and continuous involvement and coordination.
- MOUs and PAs: developing MOUs or PAs among agencies for supporting early and continuous involvement and coordination.
- Training and Outreach: providing agencies with a common understanding of one another's roles and responsibilities through webinars and group meetings.

3.3.3 Performance Measures

A set of performance measures for evaluating the future implementation of the integrated process were identified based on a review of the different performance measures used in other states (e.g., FDOT 2005) and recommendations from unstructured meetings with eight experts from IDOT, FHWA, and MPOs. Two main types of performance measures were identified: interagency coordination and communication performance measures and project delivery performance measures. A sample of the performance measures is shown in Table 3.5.

Table 3.5 – A Sample of Performance Measures

Performance measure	Information source
The percentage of interagency advisory group reviews completed within the defined review period, during the planning screens	Planning screen summary reports
The percentage of interagency advisory group reviews completed within the defined review period, during the programming screens	Programming screen summary reports
The average length of Environmental Assessment (EA) processing time	Project And Program Action Information System (PAPAI)
The average length of Environmental Impact Statement (EIS) processing time	Project And Program Action Information System (PAPAI)

3.4 Validating the Integrated IDOT-MPO-NEPA Planning Process

A second expert survey was conducted for validation. The purpose of the survey was to validate the integrated process and evaluate its specific implementation details.

3.4.1 Questionnaire Design

A validation questionnaire was used for soliciting expert opinion. The questionnaire was composed of five sections: (1) respondent information, (2) collaboration-oriented practices, (3) process-oriented practices, (4) process representation and interaction, and (5) integrated process performance measures.

Section1 aimed to collect the following respondent information: name, contact information, agency he/she represents, and years of experience. Section 2 and Section 3 aimed to evaluate the specific implementation details of the collaboration-oriented and process-oriented practices, respectively. For example, experts were asked whether they agree with the composition of the interagency advisory group, as described in the guidance document. Section 4 aimed to evaluate the process representation and interactions. For example, experts were asked whether they agree with the inputs and outputs of each of the subprocesses of the integrated process, as described in the guidance document. Section 5 aimed to evaluate the performance measures for assessing the implementation of the integrated process. An open-ended question was also provided after each question to allow experts to add suggestions or recommendations. Section 2 to Section 5 included a total of 34 questions, and similar to the first survey, a six-point Likert scale was used to record the responses of respondents, with six being the most favorable.

3.4.2 Verifying the Questionnaire Design

Before proceeding with the validation survey, a pilot study was conducted with eight experts from IDOT, FHWA, and MPOs to evaluate the effectiveness of the questionnaire. Similar to the first pilot study, the experts were requested to review the questionnaire and then to provide feedback on its format and content. The questionnaire was revised according to the feedback. For example, open-ended questions were added to allow experts to suggest adding any performance measures or deleting any inappropriate and/or irrelevant ones.

3.4.3 Survey Implementation

To solicit expert feedback in an efficient manner, a draft guidance document describing the integrated process and the questionnaire were sent to each of the interviewees two weeks prior to the interview date to allow interviewees sufficient time for review. Each interview consisted of two parts. The first part covered a detailed presentation about the integrated process. In the second part of the interview, the interviewees were asked to complete the questionnaire to gather their opinions on the proposed integrated process. The validation survey targeted the same four groups of experts, and was also conducted in a one-on-one interview format. A total of thirteen experts (including seven experts who participated the first survey) were interviewed: four from IDOT districts, four from MPOs, three from resource agencies, and two from IDOT Central Office and FHWA. In this survey, all interviews were conducted face-to-face because of the high level of detail involved.

3.4.4 Survey Results

The mean, standard deviation, and median scores were calculated, and the results were interpreted based on the median scores. A sample of the survey results are summarized in Table 3.6. The results indicate that collectively all thirteen experts "agree" with all implementation details of the collaboration-oriented and process-oriented practices, with the process representation and interactions, and with all performance measures.

Table 3.6 – A Sample of Survey Results of the Validation of the Integrated IDOT-MPO-NEPA Process

Implementation details	Mean	Standard deviation	Median	Overall opinion of respondents (based on median)
Implementation details of the collaboration-oriented integration practices				
Functions of the common database	5.63	0.52	5	Agree
Implementation details of the process-oriented integration practices				
Procedure for interagency coordination during the development of the MPO's LRTP	4.63	0.83	5	Agree
Representation and interaction of the subprocesses				
Process interactions shown in the IDOT-MPO-NEPA Integrated Planning Process Flowchart	5.25	0.44	5	Agree
Performance measures for evaluation of the Integrated IDOT-MPO-NEPA Planning Process				
Interagency coordination and communication performance measures	4.75	0.45	5	Agree

3.4.5 Validating the Reliability of the Survey Results

A Cronbach's alpha test was also conducted to validate the reliability of the survey results. The overall Cronbach's alpha value for the validation survey is 0.93, which indicates a high level of reliability.

CHAPTER 4: SEMANTIC ANNOTATION FOR CONTEXT-AWARE INFORMATION RETRIEVAL

4.1 State of the Art and Knowledge Gaps

In recent years, a number of information retrieval systems, in the computer science (CS) domain, have been developed using ontology-based semantic annotation (SA) methodologies to help users better clarify their information needs. Depending on the usage of the ontology and the level of semantic analysis, these ontology-based SA methodologies can be classified into two primary categories: shallow SA and deep SA approaches. Shallow SA approaches use mainly syntactic features to annotate the text with the concepts in the ontology. For example, Kiryakov et al. (2004) utilized pre-populated lexical resources, such as an organization name thesaurus, to annotate documents with the concepts from the knowledge and information management (KIM) ontology (Popov et al. 2003). Deep SA approaches, in contrast, use mainly semantic features to annotate the text with the concepts in the ontology. For example, Fernandez et al. (2011) adopted a scalable approach to annotate documents based on the statistical occurrences of semantic entities and their contextual semantic information. To improve the accuracy of SA, few researchers (Fernandez et al. 2011, Nesic et al. 2010) have used a combination of shallow and deep approaches. For example, Nesic et al. (2010) applied the lexical expansion of concept descriptions to calculate the weight of each syntactic match, and used the concept exploration algorithms to discover relevant semantic matches and calculate semantic distances between syntactic and semantic matches.

Although extensive studies on SA have been conducted in the CS domain, there still exist many challenges in developing a semantic annotator that can efficiently generate substantial amount of accurate semantic annotations, which is central to the implementation of a context-aware information retrieval system (Nesic et al. 2010): (1) The performance of SA could be negatively affected by inaccurate descriptions of ontology concepts and/or possible ambiguities in the meaning of concept labels (Nesic et al. 2010); (2) Because not all concepts are equally relevant to the resource they annotate, it is important to evaluate annotation relevance of the discovered concepts for the purpose of selecting the most relevant ones; (3) Most of the current SA methodologies have not been evaluated on domain-specific ontologies, which typically have more sparse concept space and more complex concept relationships. For example, Kiryakov et al. (2004) only applied their SA algorithms on a light-weight upper-level ontology, and Nesic et al. (2010) developed their SA algorithms based on the KIM knowledge base (Popov et al. 2003), which is an ontology that covers knowledge of general importance such as geographic locations and organizations; and (4) The performance of shallow and deep SA approaches have not been compared comprehensively. For example, Kiryakov et al. (2004) only investigated the shallow SA approach, and Fernandez et al. (2010) applied both shallow and deep SA approaches but did not compare the performance of the two.

4.2 Proposed Semantic Annotation Method

To address the aforementioned gaps and needs, this research task explores, both, shallow and deep SA approaches. For shallow SA, (1) In order to improve the performance of SA, syntactic concept expansion was conducted to expand concept term(s) with syntactically-related terms, syntactic concept filtering was conducted to remove noise brought by syntactic concept expansion, and domain-specific concept expansion was performed to expand concept term(s) with domain-specific context terms; and (2) Term frequency-inverse document frequency (TF-IDF) weighting and lexical relations between the original concept term and the expansion concept terms were used to determine the relevance of SA. For deep SA, (1) In order to improve the performance of SA, semantic concept expansion was conducted to expand concept term(s) with terms from semantically related concepts; and (2) TF-IDF weighting and semantic similarities between the original concept and the expansion concepts were used to determine the relevance of SA. Both the shallow and deep approaches utilized the TPER epistemology, which is a domain-specific semantic model. The performance of the two approaches were evaluated and compared on a testing data set of 1,328 Web pages.

The methodology for SA for supporting context-aware information retrieval in the TPER domain is, thus, composed of six main steps, as per Figure 4.1: (1) step 1: TPER epistemology development, (2) step 2: data preparation, (3) step 3: data preprocessing, (4) step 4: shallow SA, (5) step 5: deep SA, and (6) step 6: evaluation. In steps 4 and 5 shallow and deep SA algorithms are proposed, respectively, as alternative ways for conducting SA. Step 6 is conducted to evaluate the proposed shallow and deep SA algorithms based on performance and accordingly select the final proposed SA algorithm.

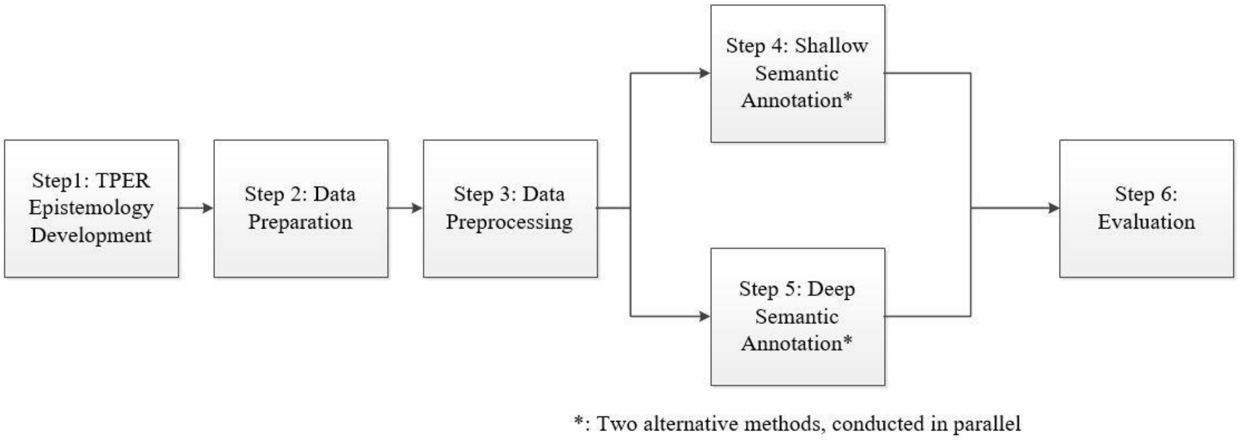

Step1: TPER Epistemology Development
Step 2: Data Preparation
Step 3: Data Preprocessing
Step 4: Shallow Semantic Annotation*
Step 5: Deep Semantic Annotation*
Step 6: Evaluation
*: Two alternative methods, conducted in parallel

Figure 4.1 – Semantic Annotation Methodology

4.2.1 TPER Epistemology Development

The TPER epistemology aims to support context-aware, domain-specific information retrieval through modelling the context dimensions of information and information retrieval in the TPER domain. In developing the TPER epistemology, the context-aware epistemic model for sustainable construction practices by Zhang and El-Gohary (2015) was benchmarked. The concepts in the epistemology were defined based on a literature review of work in the following three subdomains: (1) epistemology and its application in different domains (e.g., Muis 2004, Honderich 1995, Steup 2011, Alavi and Leidner 2001, De Jong 1996), (2) context-aware information retrieval systems (e.g., Bahrami 2007, Fernandez et al. 2011, Nesic et al. 2010, Ozcan and Aslandogan 2005), and (3) transportation project environmental review (e.g., Barberio et al. 2008a, CEQ 2007, FHWA 2011, ICT 2014, IDOT 2010). The most abstract concept in the TEPR epistemology is the "TPER epistemic context", which includes "user context", "searching context", and "document context". A TPER epistemic context describes the set of circumstances, situations, settings, environments, characteristics, or parameters that influence and/or characterize the user, the process of searching for relevant documents, and the documents in the TPER domain. A user context describes the set of characteristics and settings of the user who conducts the searching process, in terms of specific interests, preferences, task(s) at hand, and/or personal profile of the user. A searching context describes the set of circumstances, situations, settings, and/or environments in which a searching process occurs, in terms of the searching device, searching source, searching method, and searching environment. A document context describes a collection of relevant conditions and settings that make the semantics of the document unique and comprehensible to that condition, such as project context, functional process context, and resource context. A partial view of the TPER epistemology is shown in Figure 4.2.

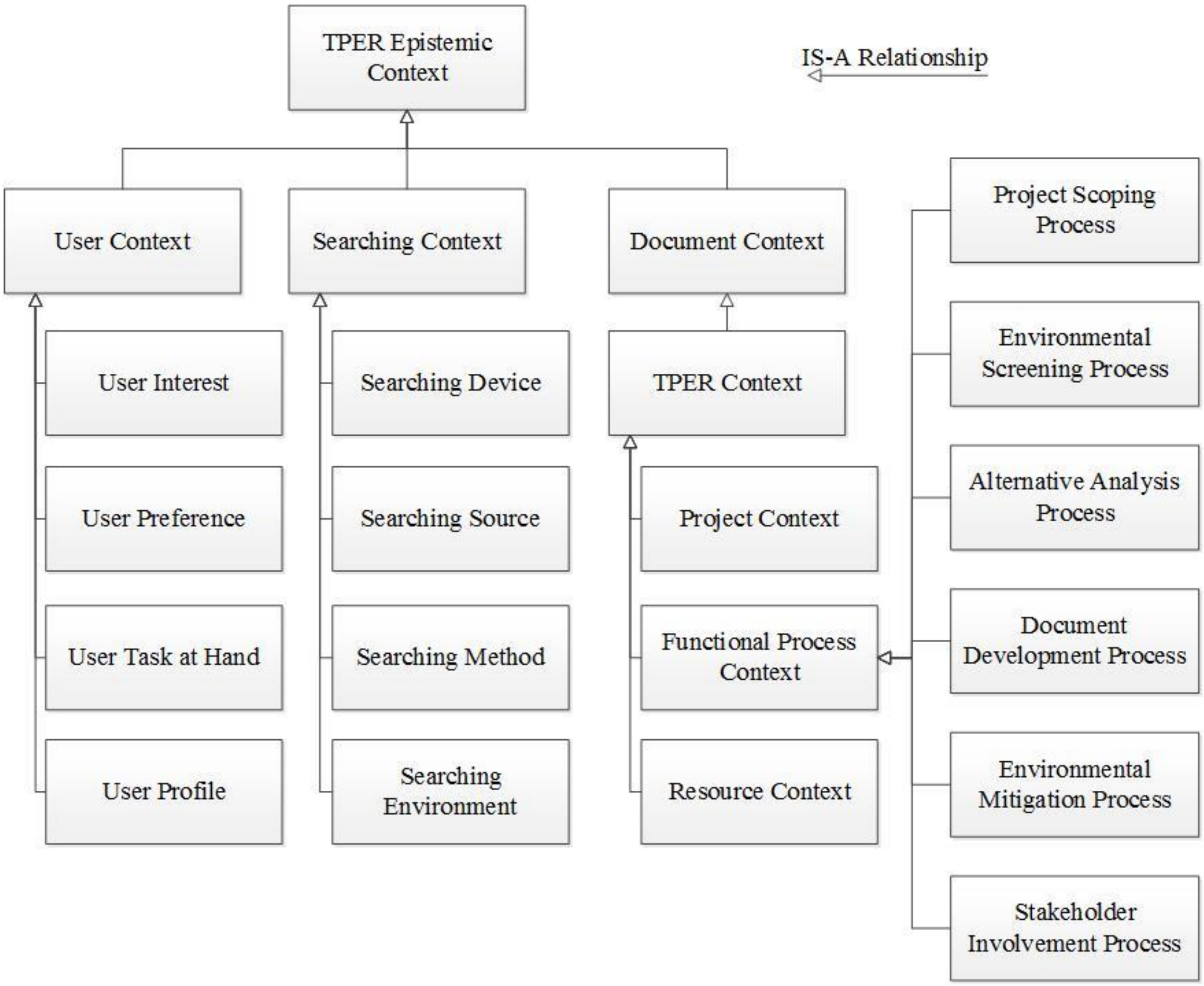


Figure 4.2 – Partial View of the TPER Epistemology

4.2.2 Data Preparation

Data preparation included two main steps: (1) data collection, and (2) manual annotation for developing the gold standard. To create a document collection for testing the proposed SA methods in the TPER domain, around 3,300 Web pages were crawled under the domain of the FHWA Environmental Review Toolkit website (www.environment.fhwa.dot.gov). The FHWA Environmental Review Toolkit home page (http://www.environment.fhwa.dot.gov/index.asp) was selected as the seed page. Starting from the seed page, every Web page under the domain was examined and its URL, title, and textual content of the body part were stored in a .txt format local file using Scrapy (http://scrapy.org/), a python-based Web crawler. When writing the crawled information into the local file, their encodings were automatically transformed into UTF-8 encoding, and any html tags (such as <head>) and non-ASCII characters (such as Spanish words) were removed to ensure that the performance of concept matching is not undermined by noise that is irrelevant to the content of the document. A document in the collection consists of the title and the textual content of the body part of the respective crawled Web page. Web pages that do not have textual contents in their body parts (e.g., only have images or videos) were further excluded from the document collection. Web pages that have textual contents in their body parts but were redundant were also excluded. Accordingly, the final document collection contains 1,328 Web pages.

After data collection, each document was manually annotated by three annotators (the author and two other researchers) with one or more functional process context concepts. This thesis focuses on analyzing the "functional process context", a subconcept of the "TPER context". As per Figure 4.2, the "functional process context" has 6 subconcepts: "project scoping process", "environmental screening process", "alternative analysis process", "document development process", "environmental mitigation process", and "stakeholder involvement process". Each annotator independently annotated each document with zero or more functional process context concepts. For each document, the annotation was based on the agreement between annotators. Two main methods were used for discrepancy resolution: (1) If a majority (i.e., at least two) of annotators achieved agreement, then the agreed-on annotation was used; and (2) If a majority of annotators did not achieve agreement, then a discussion was conducted until a majority agreement was achieved. The 1,328 documents were annotated with a total of 2,958 concepts, with an average of

2.23 annotation concepts per document. The manual annotation results formed the gold standard for the following experiments.

4.2.3 Data Preprocessing

To prepare the raw text data for the implementation of SA algorithms, the bag of words (BOW) model was used to represent each document. In this model, a document is represented as an unsorted set of words with their corresponding weight that represents the discriminating power of the word. As the most commonly-used weighting scheme in information retrieval problems, TF-IDF weighting was adopted for conducting SA. In order to represent a document using the BOW model, the following three techniques for data preprocessing were conducted: (1) Tokenization: Tokenization is the process of breaking the text into tokens, which are meaningful elements such as words, phrases, or symbols. The tokenization process removes certain characters like punctuations and transforms the words into their lowercase forms. In this work, a single word was regarded as a common token, and a list of special (domain-specific) tokens that consist of terminologies in the TEPR domain was also developed. Examples of these special tokens include "categorical exclusion", "environmental assessment", and "environmental impact statement", which refer to the three different environment review actions required by the federal law; (2) Stopword removal: Stopwords are those words that have high frequency but low discriminating power, which have little value in helping select documents that match a user need. Removing stopwords can, thus, help eliminate nondiscriminative high-frequency words, thereby reducing the number of features and revealing the discriminative words; and (3) Lemmatization: Lemmatization is the process of removing inflectional endings and returning the base or dictionary form of a word, which is known as the lemma. By combining words with the same lemma, lemmatization can reduce the number of features, and can be effective in enhancing the performance of SA. For example, after the lemmatization, the words “mitigates”, “mitigated”, and “mitigating” would all be transformed into their lemma “mitigate”.

4.2.4 Shallow Semantic Annotation

The proposed shallow SA approach uses syntactic features to annotate the text with the concepts in the TPER epistemology. For each concept (e.g., “stakeholder involvement process”) in the TPER epistemology, a concept index was created to store the concept terms (e.g., “stakeholder” and “involvement”) of the concept, which are the most common text descriptions of the concept. Syntactic concept expansion was then performed to expand each concept term with its related lexical terms from a lexical dictionary. Syntactic concept filtering was subsequently conducted to filter the noise that was introduced as a result of concept expansion. Further, each concept term was expanded with related domain-specific terms. The relevance of the annotation was then determined by the TF-IDF weights of the original concept terms and expansion concept terms and the relations between the expansion concept terms and original concept terms. The shallow SA approach is, thus, performed in three steps: (1) syntactic concept expansion, (2) syntactic concept filtering and domain-specific concept expansion, and (3) syntactic terms matching.

4.2.4.1 Syntactic Concept Expansion

Syntactic concept expansion, in this research, aims to expand the concept index of each concept in the TPER epistemology with related terms from the WordNet, a lexical dictionary. Three types of semantic relations were considered: synonyms, hyponyms, and hypernyms. Synonyms are the terms that share the same meaning. A hypernym is a term that describes a broader semantic category than that of another term. A hyponym is a term that refers to a more specific semantic category than that of another term. A hypernym-hyponym relationship, thus, reflects a superconcept-subconcept relationship. For example, the term “screen” was selected as the description for the concept “environmental screening process” and, accordingly, the concept index after syntactic concept expansion is shown in Table 4.1.

Table 4.1 – Example of Concept Index after Syntactic Concept Expansion (Partial Concept Index of “Environmental Screening Process”)

Type of terms	Terms in concept index
Synonyms	'blind', 'screenland', 'shield', 'covert', 'cover', 'riddle', 'sort', 'test', 'filmdom', 'sieve', 'concealment'
Hyponyms	'blind', 'sifter', 'mantle', 'blinder', 'surface', 'obturate', 'strain', 'examine', 'smokescreen', 'pall', 'canvass', 'door', 'winker', 'windshield', 'check', 'select', 'canvas', 'show', 'camouflage', 'desktop', 'jam', 'choose', 'shoji', 'curtain', 'analyze', 'divider', 'drapery', 'take', 'blinker', 'purdah', 'fireguard', 'analyse', 'reredos', 'sift', 'protection', 'impede', 'background', 'strainer', 'sieve', 'windscreen', 'stalking-horse', 'protect', 'drape', 'covering', 'altarpiece', 'study', 'partition', 'riddle', 'shutter', 'occlude', 'shade', 'obstruct', 'display', 'block'
Hypernyms	-

4.2.4.2 Syntactic Concept Filtering and Domain-Specific Concept Expansion

Syntactic concept filtering and domain-specific concept expansion, in this research, aims to (1) remove the noise brought by syntactic concept expansion, and (2) expand concept terms with domain-specific context terms. As a database mainly built on lexical analysis, WordNet covers a limited number of semantic relations and is independent of any document collection. Because of these two characteristics, WordNet (1) cannot expand a term with domain-specific semantic relations, and (2) may bring noise to the expansion, which would undermine the performance of syntactic matching (Gong et al. 2006). For example, as shown in Table 4.1, the expanded concept index includes a lot of terms that are not relevant to the concept “environmental screening process”, such as “blind”, “mantle”, and “door”. To overcome this limitation of WordNet, term association rules were applied to filter noisy terms and add domain-specific semantically related terms as a supplement.

Term associations were assessed using confidence and support. Confidence is the conditional probability that a document that contains term t_i also contains term t_j (Han et al. 2011). Support is the probability that a document contains both term t_i and t_j (Han et al. 2011). Confidence and support were calculated using Eq. (4.1) and Eq. (4.2) [Tan et al. (2013)], respectively, where the $D\left(t_i, t_j\right)$ is the total number of documents that contain both concept terms t_i and t_j, $D(t_i)$ is the total number of documents that contain concept t_i, and the *N* is the total number of documents in the collection.

$$Conf_{t_i - t_j} = P(\, t_j | \, t_i \,) \; = \frac{D(t_i, t_j)}{D(t_i)} \tag{4.1}$$

$$Sup_{t_i - t_j} = P\left(t_i \cup t_j\,\right) = \; \frac{D(t_i, t_j)}{N} \tag{4.2}$$

To remove/control the noise, a threshold of confidence and support was used. A high threshold could remove useful expansion terms that are meaningful to the original concept, and a low threshold may keep noisy expansion terms that are not meaningful to the original concept. A range of values (confidence from 0 to 1 with intervals of 0.1, and support from 0 to 0.1 with intervals of 0.01) were tested to empirically find the optimized threshold values; the testing results were evaluated and the values that yielded the highest performance were selected. Based on the empirical results, confidence and support threshold values of 0.3 and 0.01, respectively, were used. Accordingly, only candidate expansion terms that have a confidence over 0.3 and a support over 0.01 with the original concept term were added to the concept index.

To expand an original concept term with domain-specific context terms, the top candidate context terms that have the highest confidence and support with the original concept term were selected as the context terms. A set of values ranging from 1 to 20 with intervals of 1 were empirically tested to find the optimal number of context terms to use for expansion. Based on the empirical results, ten domain-specific context terms were used. During the selection of context terms, domain-specific stopwords that have little discriminating power in the domain were disregarded. A term is considered as a domain-specific stopword, if it appears in over a threshold of the documents. A set of threshold values ranging from 10% to 100% with intervals of 10% were tested. Based on the empirical results, a threshold value of 50% was used. The disregarded domain-specific stopwords include terms such as "process", "project", "FHWA", and "transportation". After the domain-specific stopwords were removed, the context terms were added to the concept index. For example, the context terms for "screen" are "develop", "identify", "work", "design", "make", "study", "impact", "base", "level", "exist", and "help". Most of these terms are not described in the WordNet expansion.

4.2.4.3 Syntactic Term Matching

Syntactic term matching, in this study, aims to calculate the relevance of the annotation of the concepts in the TPER epistemology. After concept expansion and concept filtering, a concept c_i in the TPER epistemology has a concept index (CI_i), as reflected in Eq. (4.3), where $t_j$is an original concept term (from the TPER epistemology) for concept c_i, m is the total number of original concept terms for concept c_j, t_{jk} is an expansion concept term (synonyms, hyponyms, hypernyms, or context term) for an original concept term t_j, and n is the total number of expansion concept terms for an original concept term t_j for a concept c_i .

$$CI_i = \{\cup_{j=1}^{m} t_j\, , \cup_{j=1}^{m} \cup_{k=1}^{n} t_{jk}\} \tag{4.3}$$

For a set of expansion concept terms, each expansion term could have a different semantic relevance to the original concept term. A term relevance factor (TRF) was, thus, proposed in this work to differentiate the degrees of relevance of expansion concept terms to an original concept term. The concept relevance vector, $\vec{R}(CI_i)$, is expressed in Eq. (4.4), where $TRF(t_j)$ is the TRF of t_j, which is an original concept term of concept c_i, and $TRF(t_{jk})$ is the TRF of t_{jk} , which is an expansion concept term of concept c_i. The values of $TRF(t_j)$ are equal to 1. For an expansion concept term t_{jk} whose original concept term is t_j , $TRF(t_{jk})$ has a value of δ_{syn} if t_{jk} is a synonym of t_j, has a value of δ_{hypo} if t_{jk} is a hyponym of t_j, has a value of δ_{hyper} if t_{jk} is a hypermym of t_j , and has a value of $\delta_{context}$ if t_{jk} is a context term of t_j. In order to optimize the performance of syntactic matching, a range of values (from 0 to 1 with intervals of 0.01) were tested for δ_{syn}, δ_{hypo}, δ_{hyper}, and $\delta_{context}$. Based on the best performance results, δ_{syn} = 0.26, δ_{hypo}= 0.28, δ_{hyper} = 0, and $\delta_{context}$ = 0.46 were used for the experiments conducted in this research.

$$\vec{R}(CI_i) = \{\cup_{j=1}^{m} TRF(t_j) , \cup_{j=1}^{m} \cup_{k=1}^{n} TRF(t_{jk})\} \tag{4.4}$$

The documents were then automatically searched to check if the concept terms in the concept index appear in these documents. For the concepts whose terms appear in the documents, the annotation weight of each concept was calculated by considering the following two factors: (1) the TF-IDF weights of the concept terms in the document, and (2) the concept's $\vec{R}(CI_i)$. For document d and concept term t_j of concept c_i, the TF-IDF weight $W_{t_j}$was defined using Eq. (4.5) (Manning et al. 2009), where $TF(t_j, d)$ is the frequency of term t_j in document d, and $IDF(t_j)$ is the inverse document frequency of term t_j in the document collection. The inverse document frequency $IDF(t_j)$ was calculated using Eq. (4.6) (Manning et al. 2009), where N is the total number of documents in the collection, and $||D(t_j)||$ is the number of documents that include term t_j.

$$W_{t_j} = \log(TF(t_j, d) + 1) * [1+ \log (IDF(t_j))] \tag{4.5}$$

$$IDF(t_j) = \frac{N}{||D(t_j)||} \tag{4.6}$$

For document d and concept c_i, the concept weight vector $\overrightarrow{W}(C_i|d)$ was defined using Eq. (4.7), where W_{t_j} is the TF-IDF weight of concept term t_j of concept c_i , and W_{jk} is the TF-IDF weight of expansion concept term t_{jk} of concept c_i. For a document d and concept c_i, the annotation weight $W_{c_i}(d)$ was defined using Eq. (4.8) (Nesic et al. 2010), where $\overrightarrow{W}(C_i|d)$ is the concept weight vector of concept c_i, and $\vec{R}(CI_i)$ is the concept relevance vector of concept c_i.

$$\overrightarrow{W}(C_i|d) = \{\cup_{j=1}^{m} W_{t_j}, \cup_{j=1}^{m} \cup_{k=1}^{n} W_{t_{jk}}\} \tag{4.7}$$

$$W_{c_i}(d) = \overrightarrow{W}(C_i|d) * \vec{R}(CI_i) \tag{4.8}$$

4.2.5 Deep Semantic Annotation

The proposed deep SA approach (1) uses the TPER epistemology as one of the inputs for annotation, and (2) involves deep semantic analysis. For each concept in the TPER epistemology, a semantic concept index was created to not only contain the concept terms of the original concept, but also the concept terms of its related concepts. The relevance of the annotation was then determined by the TF-IDF weights of the terms in the semantic concept index and the semantic similarities between the original concept and the related concepts. The deep SA approach is, thus, performed in three steps: (1) semantic concept expansion, (2) semantic similarity assessment, and (3) semantic term matching.

4.2.5.1 Semantic Concept Expansion

Semantic concept expansion aims to expand the concept indexes of the concepts in the TPER epistemology with terms from its semantically related concepts. For a concept c_i in the TPER epistemology, its semantically related concepts include its descendants (direct and indirect subconcepts) and other concepts that have non-hierarchical relations to concept c_i. For example, as shown in Figure 4.3, the semantically related concepts of concept “environmental screening process” include “environmental resource”, “impact analysis”, “data collection process”, “gis analysis process”, “database collection”, “field collection”, and “sensor collection”.

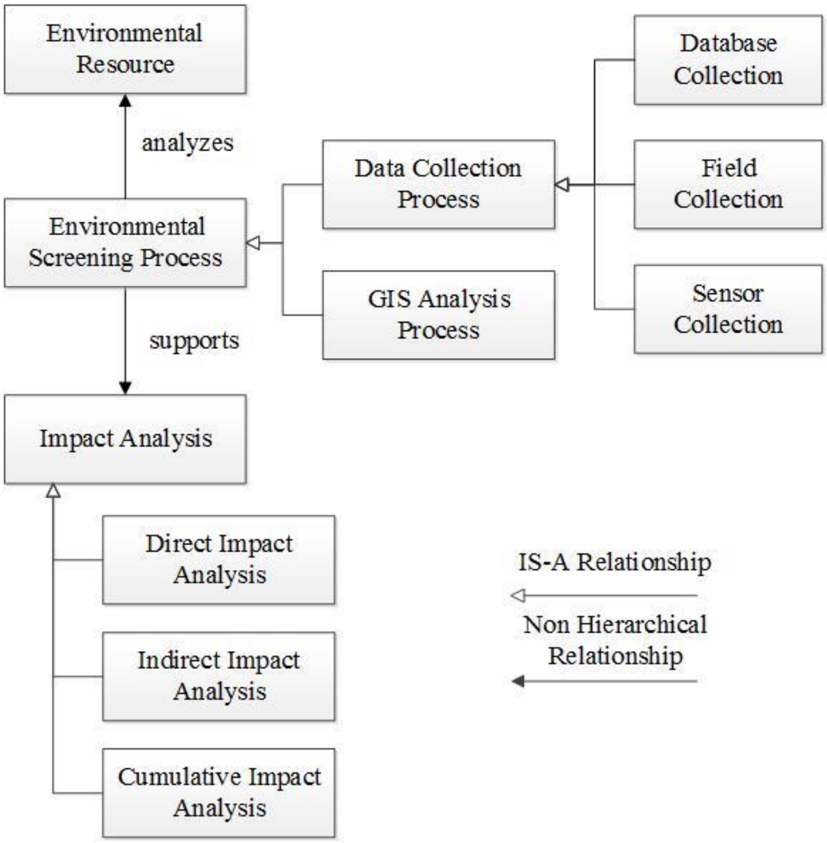


Figure 4.3 – Related Concepts for “Environmental Screening Process”

The semantic concept index (CI_i^s) of concept c_i after semantic concept expansion is shown in Eq. (4.9), where t_j is an original concept term of concept c_i, m is the total number of original concept terms for concept c_i, t_{lk} is a concept term for concept c_l that acts as an expansion concept term for concept c_i, c_l is a concept that is semantically related to concept c_i, p is the total number of semantically related concepts for concept c_i, and q is the total number of concept terms for concept c_l.

$$CI_i^s = \{\bigcup_{j=1}^{m} t_j\,, \bigcup_{l=1}^{p} \bigcup_{k=1}^{q} t_{lk}\} \tag{4.9}$$

Similar to syntactic concept expansion, TRF was used to differentiate the degrees of relevance of expansion concept terms to an original concept term. The concept relevance vector, $\vec{R}(CI_i^s)$, of concept c_i is shown in Eq. (4.10), where $TRF(t_j)$ is the TRF of t_j, which is an original concept term of concept c_i; and $TRF(t_{lk})$ is the TRF of t_{lk}, which is a semantic expansion concept term of concept c_i. The values of $TRF(t_j)$ are equal to 1. As per Eq. (4.11), for an expansion concept term t_{lk} which belongs to concept c_l, $TRF(t_{lk})$ has a value of $S_n(c_i, c_l)$, which is the semantic similarity between the original concept c_i and its semantically related concept c_l.

$$\vec{R}(CI_i^s) = \{\bigcup_{j=1}^{m} TRF(t_j)\,, \bigcup_{l=1}^{p} \bigcup_{k=1}^{q} TRF(t_{lk})\} \tag{4.10}$$

$$\vec{R}(CI_i^s) = \{\bigcup_{j=1}^{m} TRF(t_j)\,, \bigcup_{l=1}^{p} \bigcup_{k=1}^{q} S_n(c_i, c_l)\} \tag{4.11}$$

4.2.5.2 Semantic Similarity Assessment

In this research, eight SS measures were tested and evaluated. These measures are typically used to assess the SS between pairs of concepts based on a general domain ontology such as the KIM ontology (Popov et al. 2003). The use of the TPER epistemology, as opposed to other general domain ontologies, allows for enhanced SS assessment because SS is assessed based on domain knowledge. The eight tested SS measures (which are classified into path-based, node-based, and combined) are: Wu and Palmer (1994) (path-based), Leacock and Chodorow (1998) (path-based), Li et al. (2003) (path-based), Mao and Chu(2007) (path-based), Resnik (1995) (node-based), Jiang and Conrath (1997) (node-based), Lin(1998) (node-based), and Al-Mubaid and Nguyen (2006) (combined approach).

To achieve the best performance in SA, parameter tuning was conducted for Li et al. (2003) SS, Mao and Chu (2007) SS, and Al-Mubaid and Nguyen (2006) SS. For Li et al. (2003) SS, a set of values (0 to 1 with intervals of 0.1) for the scaling factors α and β were tested, and accordingly α and β were set to 0.2 and 0.6, respectively, based on the performance results. For Mao and Chu (2007) SS, a range of values (0 to 1 with intervals of 0.1) for the upper-bound similarity value δ were tested, and accordingly a value of 0.9 was selected. For Al-Mubaid and Nguyen (2006) SS, (1) The k value was set to 1 to ensure that the semantic distance between two same concepts is 0; and (2) A range of values (1 to 5 with intervals of 1) for the scaling factors α and β were tested, and accordingly α and β were set to 3 and 1, respectively.

4.2.5.3 Semantic Term Matching

In order to ensure that all the different SS measures are in a notionally common scale, a min-max normalization was conducted, as per Eq. (4.12), where $S_n(c_i, c_l)$ is the normalized SS score between concepts c_i and c_l, $S(c_i, c_l)$ is the calculated SS score based on an SS measure, $S_{min}(c_i, c_l)$ is the minimal SS score between any two concepts in the hierarchy, and $S_{max}(c_i, c_l)$ is the maximal SS score between any two concepts in the hierarchy. Accordingly, $S_n(c_i, c_l)$ ranges from 0 to 1, where S_n is equal to 0 when concepts c_i and c_l are the least similar concepts in the hierarchy, and S_n is equal to 1 when concepts c_i and c_l are the same concepts.

$$S_n(c_i, c_l) = \frac{S(c_i,c_l) - S_{min}(c_i,c_l)}{S_{max}(c_i,c_l) - S_{min}(c_i,c_l)} \quad (4.12)$$

After SS score normalization, the concept relevance vectors were expressed, as per Eq. (4.11). Similar to shallow SA, the documents were then automatically searched to check if the concept terms in the semantic concept index appear in these documents. For the concepts whose terms appear in the documents, the annotation weights of the concept were calculated – as per Eq. (4.5) to Eq. (4.8) – based on (1) the TF-IDF weights of the concept terms in the document, and (2) the concept's $\vec{R}(CI_i)$.

4.2.6 Evaluation

For a concept c_i, each document in the collection was ranked based on the annotation weight $W_{c_i}(d)$ and then the documents with top *w* annotation weights were annotated as relevant to the concept c_i. The performances of shallow and deep SA were evaluated using mean precision (MP) and mean average precision (MAP) at the top *k* documents.

For a concept c_i, precision was calculated based on Eq. (4.13), where true positive (*TP*) refers to the number of documents annotated correctly and false positive (*FP*) refers to the number of documents annotated incorrectly. MP for a set of concepts is the arithmetic mean of the precision values of the concepts. MP was calculated as per Eq. (4.14), where $P(c_i)$ is the precision of concept c_i and *B* is the total number of concepts. MAP was calculated based on precision and average precision (AP). AP is the average precision values at the ranks where correctly annotated documents occur (i.e., at the ranks where recall changes). As such, AP provides a single measure that evaluates the combined performance of precision, recall, and the ranking order.

AP was calculated as per Eq. (4.15), where *k* is the rank of document based on the annotation weight $W_{c_i}(d)$ for concept c_i, *A* is the total number of annotated documents, *P(k)* is the precision value at rank *k*, and *rel(k)* is an indicator function equals to 1 if the annotated document at rank k is annotated correctly and 0 otherwise. MAP for a set of concepts is the arithmetic mean of the AP values of the concepts. Accordingly, MAP was calculated as per Eq. (4.16), where *AP(c_i)* is the average precision of concept c_i and *B* is the total number of concepts. For the experiments conducted in this research, MP and MAP values at the top 10, 20, 30, 40, and 50 annotated documents were calculated.

$$P(c_i) = \frac{TP}{TP+FP} \tag{4.13}$$

$$MP = \frac{P(c_i)}{B} \quad (4.14)$$

$$AP(c_i) = \frac{\sum_{k=1}^{A} P(k) * rel(k)}{A} \quad (4.15)$$

$$MAP = \frac{\sum_{i=1}^{B} AP(c_i)}{B} \quad (4.16)$$

4.3 Experimental Results and Analysis

4.3.1 Performance of Shallow Semantic Annotation

Shallow SA was conducted in three different ways: (1) using original concept terms only, (2) conducting concept expansion on original concept terms, and (3) conducting both concept expansion and filtering on original concept terms. The performance results of the three shallow SA methods are summarized in Tables 4.2, 4.3, and 4.4.

As shown in Table 4.2, when using only original concept terms, (1) at the top 10 to 50 documents MP values were 53%, 58%, 62%, 67%, and 70%, respectively, and MAP values were 35%, 37%, 40%, 44%, and 47%, respectively, and (2) the "environmental mitigation process" concept consistently achieved a best performance of 100%. The incorrect annotations in this case are largely due to the ambiguity, or double meanings, of concept descriptions (original concept terms). For example, the original concept term "alternative" of the concept "alternative analysis process" could act as a noun (in the TPER domain, that usually refers to concepts like "project alternative" or "design alternative") or adjective (in the TPER domain, that usually refers to concepts like "alternative fuel" or "alternative transportation"). The perfect performance shown for the "environmental mitigation process", on the other hand, is likely due to the lower ambiguity, or standardized meanings, of concept descriptions. For example, the original concept term "mitigation" of the concept "environmental mitigation process" is much less ambiguous in the TPER domain; it usually refers to the concept "environmental mitigation measure".

Table 4.2 – Performance of Shallow Semantic Annotation Using Original Concept Terms Only

	Precision (P) at top k					Average precision (AP) at top k				
Concept	P at top 10	P at top 20	P at top 30	P at top 40	P at top 50	AP at top 10	AP at top 20	AP at top 30	AP at top 40	AP at top 50
PSP*	40%	45%	50%	50%	52%	14%	19%	22%	23%	25%
ESP*	50%	45%	53%	65%	70%	19%	21%	26%	34%	40%
DDP*	30%	45%	53%	60%	64%	16%	19%	25%	30%	34%
AAP*	30%	35%	37%	45%	54%	16%	15%	15%	18%	24%
EMP*	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
SIP*	70%	75%	77%	80%	82%	43%	51%	54%	58%	61%
Mean**	53%	58%	62%	67%	70%	35%	37%	40%	44%	47%

*PSP: Project Scoping Process; ESP: Environmental Screening Process; DDP: Document Development Process; AAP: Alternative Analysis Process; EMP: Environmental Mitigation Process; SIP: Stakeholder Involvement Process
** Mean of AP is the MAP

As shown in Table 4.3, when conducting concept expansion, although MP and MAP at the top 10 to 50 documents were improved by a small percentage, there are a few cases that the annotation performance of one concept actually decreased. For example, precision and AP at the top 10 documents for concept "environmental screening process" decreased from 50 % and 19% to 40% and 14%, respectively. The reason for the performance drop is that concept expansion brought a lot of noise. For example, for the concept "environmental screening process", the expansion terms include "blind", "mantle", and "door", which are not meaningful to the original concept. The performance shown in this experiment indicates that concept expansion through WordNet can improve the performance of SA, but can also bring noise which may result in a performance drop.

Table 4.3 – Performance of Shallow Semantic Annotation after Syntactic Concept Expansion

	Precision (P) at top k					Average precision (AP) at top k				
Concept	P at top 10	P at top 20	P at top 30	P at top 40	P at top 50	AP at top 10	AP at top 20	AP at top 30	AP at top 40	AP at top 50
PSP*	20%	45%	57%	63%	60%	4%	15%	24%	30%	30%
ESP*	40%	55%	67%	73%	76%	14%	24%	35%	41%	47%
DDP*	50%	75%	70%	70%	72%	29%	48%	47%	47%	49%
AAP*	70%	60%	63%	65%	66%	53%	43%	44%	44%	44%
EMP*	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
SIP*	70%	70%	77%	75%	80%	49%	50%	56%	55%	60%
Mean**	58%	68%	72%	74%	76%	42%	47%	51%	53%	55%

*PSP: Project Scoping Process; ESP: Environmental Screening Process; DDP: Document Development Process; AAP: Alternative Analysis Process; EMP: Environmental Mitigation Process; SIP: Stakeholder Involvement Process
** Mean of AP is the MAP

As shown in Table 4.4, after concept filtering and domain-specific concept expansion, at the top 10 to 50 documents MP values improved from 58%, 68%, 72%, 74%, and 76% to 90%, 90%, 88%, 88%, and 88%, respectively; and MAP values improved from 42%, 47%, 51%, 53%, and 55% to 82%, 82%, 80%, 80%, and 80%, respectively. The enhanced performance is attributed to two main reasons. First, concept filtering removed noise brought by concept expansion. For example, expansion words that are not meaningful to the concept "environmental screening process", such as "blind", "mantle", and "door", were removed from the concept index after concept filtering. Second, domain-specific concept expansion expanded the original concept terms with domain-specific context terms. For example, after concept filtering, the concept index for the concept "environmental screening process" was expanded with domain-specific context terms such as "develop", "identify", "work", "design", and "make". The performance shown in this experiment indicates that concept filtering and domain-specific concept expansion are effective in improving the performance of SA through (1) removing noise brought from concept expansion, and (2) expanding concept terms with context terms.

Table 4.4 – Performance of Shallow Semantic Annotation after Syntactic Concept Expansion, Concept Filtering, and Domain-specific Concept Expansion

	Precision (P) at top k					Average precision (AP) at top k				
Concept	P at top 10	P at top 20	P at top 30	P at top 40	P at top 50	AP at top 10	AP at top 20	AP at top 30	AP at top 40	AP at top 50
PSP*	80%	70%	73%	75%	78%	78%	61%	59%	60%	62%
ESP*	80%	90%	87%	90%	88%	58%	72%	71%	76%	75%
DDP*	80%	90%	77%	75%	72%	58%	72%	62%	60%	57%
AAP*	100%	95%	97%	95%	96%	100%	94%	95%	93%	94%
EMP*	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
SIP*	100%	95%	97%	95%	94%	100%	92%	93%	92%	90%
Mean**	90%	90%	88%	88%	88%	82%	82%	80%	80%	80%

*PSP: Project Scoping Process; ESP: Environmental Screening Process; DDP: Document Development Process; AAP: Alternative Analysis Process; EMP: Environmental Mitigation Process; SIP: Stakeholder Involvement Process
**Mean of AP is the MAP

4.3.2 Performance of Deep Semantic Annotation

The performance of deep SA was evaluated for the eight SS measures. The performance results are summarized in Table 4.5. As shown in Table 4.5, Al-Mubaid and Nguyen (2006) SS measure achieved the best performance on every performance metric. This can be attributed to the following two reasons: (1) This measure integrates both path features (shortest path distance) and node features (IC of the MIS of the two concepts), while the other SS measures only consider one of these two types of features; and (2) This measure allows for parameter tuning to optimize the contributions of the path feature and node feature based on the types of hierarchy and application. For hierarchies with longer average distances between concepts, path-based SS measures tend to give an unreasonable low value to a concept pair with no direct hierarchical relationship, while node-based measures tend to overlook the hierarchical distance between the two concepts. For applications like SA, the contributions of the path feature and node feature can be tuned to optimize the annotation performance. For example, as shown in Figure 4.3, the related concepts of concept "environmental screening process" include "environmental resource", "impact analysis", "data collection process", "gis analysis process", "database collection", "field collection", and "sensor collection”. When using the path-based SS measures by Li et al. (2003), Wu and Palmer (1994), and Mao and Chu (2007), the SS values between concepts “environmental screening process” and “environmental resource” are only 0, 0, and 0.12, respectively. When using the node-based SS measure by Resnik (1995), the concept “data collection process” and its three subconcepts “database collection”, “field collection”, and “sensor collection” all have the same SS values with the concept “environmental screening process”. When using the node-based measures by Jiang and Conrath (1997) and Lin (1998), the SS values between concepts “environmental screening process” and “environmental resource” and concepts “environmental screening process” and “impact analysis” are 0.76 and 0.58, and 0.56 and 0.47, respectively, although “impact analysis” should be semantically more similar because of its shorter distance to “environmental screening process” in the hierarchy. Only Al-Mubaid and Nguyen (2006) SS measure, which combines both path features and node features, indicates reasonable SS values for both concept pairs (0.52 and 0.54, respectively).

Table 4.5 – Performance of Deep Semantic Annotation Using Different Semantic Similarity Measures

Semantic similarity measure	Mean precision (MP) at top k					Mean average precision (MAP) at top k				
	MP at top 10	MP at top 20	MP at top 30	MP at top 40	MP at top 50	MAP at top 10	MAP at top 20	MAP at top 30	MAP at top 40	MAP at top 50
Wu and Palmer	88%	88%	88%	90%	88%	84%	81%	80%	81%	80%
Leacock and Chodorow	92%	90%	91%	91%	91%	88%	85%	84%	85%	84%
Li et al.	83%	83%	86%	87%	86%	88%	74%	75%	77%	76%
Mao	68%	74%	76%	78%	78%	53%	57%	58%	59%	61%
Resnik	83%	88%	88%	87%	86%	77%	79%	80%	78%	77%
Jiang	90%	90%	91%	89%	89%	86%	85%	85%	83%	82%
Lin	83%	87%	86%	83%	83%	75%	77%	76%	74%	73%
Al-Mubaid and Nguyen	97%	94%	92%	93%	91%	96%	92%	89%	88%	86%

4.3.3 Comparison of Shallow and Deep Semantic Annotation

To compare the shallow and deep SA approaches, the best performing methods [conducting concept expansion, concept filtering, and domain-specific concept expansion on the original concept terms for shallow SA and using Al-Mubaid and Nguyen (2006) SS measure for deep SA] were compared. As shown in Table 4.6, for MP and MAP at the top 10 to 50 documents, the deep approach outperformed the shallow approach on every metric.

Table 4.6 – Performance of Shallow and Deep Semantic Annotation (SA)

SA method	Mean precision (MP) at top k					Mean average precision (MAP) at top k				
	MP at top 10	MP at top 20	MP at top 30	MP at top 40	MP at top 50	MAP at top 10	MAP at top 20	MAP at top 30	MAP at top 40	MAP at top 50
Shallow SA	90%	90%	88%	88%	88%	82%	82%	80%	80%	80%
Deep SA	97%	94%	92%	93%	91%	96%	92%	89%	88%	86%

The higher performance of the deep SA approach over the shallow one can be attributed to the following two reasons. First, the deep approach takes domain knowledge into consideration, whereas the shallow approach overlooked important semantic relations, such as “is-a”, and “is-part” relations. For example, Table 4.7 shows the variation in the concept index of the “environmental screening process” concept under both approaches. As per Table 4.7, important terms that describe the “environmental screening process”, such as “data”, “database”, and “field”, were not covered in the concept index when using shallow SA. Second, the shallow SA approach was purely based on lexical relations and corpus statistics, and its performance depends largely on the quality of the lexical dictionary and the corpus.

Table 4.7 – Concept Indexes of “Environmental Screening Process” under Shallow and Deep Semantic Annotation (SA)

SA method	Terms in concept index
Shallow SA	'screen', 'cover', 'surface', 'select', 'analyze', 'analyze', 'protection', 'protect', 'study', 'develop', 'identify', 'work', 'design', 'make', 'study', 'impact', 'base', 'level', 'exist', 'help'
Deep SA	'screen', 'environmental', 'resource', 'impact', 'analysis', 'data', 'collection', 'gis', 'database', 'field', 'sensor'

In terms of computational efficiency, for both shallow and deep approaches, the time to conduct SA only increases linearly with the number of documents in the collection, which makes both algorithms computationally efficient and suitable for annotating the large amount of information in the TPER domain.

4.3.4 Selection of Semantic Annotation Algorithm based on Performance

Accordingly, based on the experimental results, the following algorithm for conducting SA is proposed, as per Figure 4.4: (1) Conducting semantic concept expansion of the original concept terms: as a result, each concept (i.e., original concept) in the TPER epistemology has an associated semantic concept index that contains the concept terms of the original concept and the concept terms (i.e., semantic expansion terms) of the related concepts; (2) Conducting semantic similarity assessment: using Al-Mubaid and Nguyen (2006) SS measure to assess the semantic similarities between the related concepts and the original concept, where the normalized SS score of a semantic expansion term is used as the relevance factor of that term; (3) Conducting semantic term matching: calculating the relevance of SA (annotation weight) based on the TF-IDF weights of the semantic expansion terms and the relevance factor of the semantic expansion terms; and (4) Annotating with weight: annotating each document in the collection with the concepts along with an assignment of annotation weights. The proposed algorithm achieved over 91% and 86% MP and MAP at the top 10 to 50 documents, respectively.

Studies on the performance of state-of-the-art bibliographic search engines (such as Google Scholar) showed that MPs at the top 10 documents for most systems are between 60% and 80% (Walters 2011), which indicates that a trustworthy level of performance for an application should be within or above that range. The proposed algorithm achieved a higher performance, with a 97% MP at the top 10 documents. This indicates that the proposed algorithm would provide a reliable performance to support information retrieval in the TPER domain.

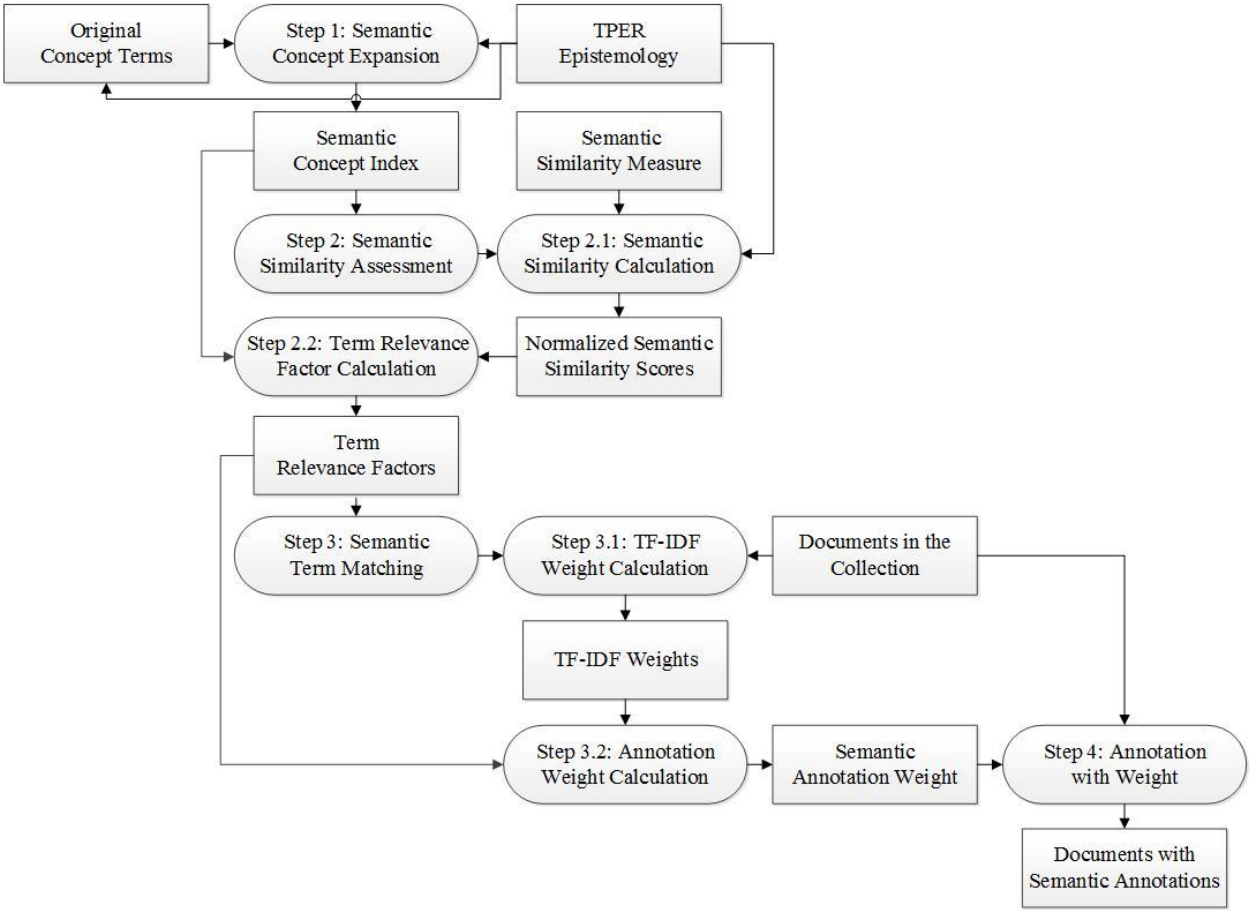
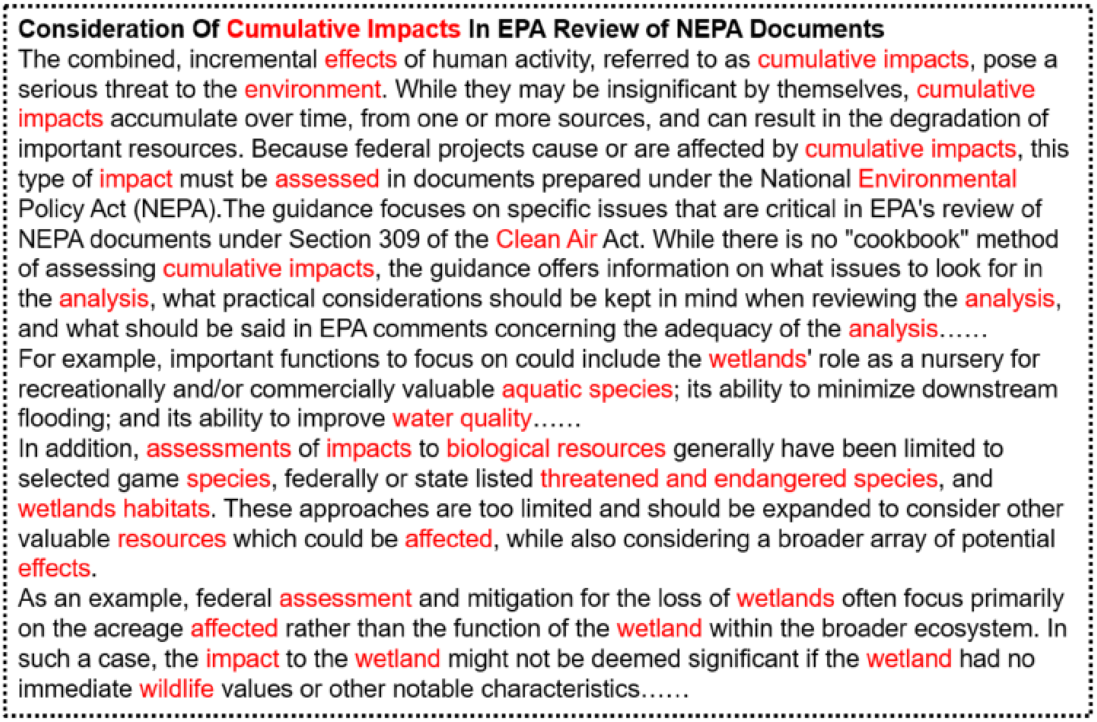

Original Concept Terms
Step 1: Semantic Concept Expansion
TPER Epistemology
Semantic Concept Index
Semantic Similarity Measure
Step 2: Semantic Similarity Assessment
Step 2.1: Semantic Similarity Calculation
Step 2.2: Term Relevance Factor Calculation
Normalized Semantic Similarity Scores
Term Relevance Factors
Step 3: Semantic Term Matching
Step 3.1: TF-IDF Weight Calculation
Documents in the Collection
TF-IDF Weights
Step 3.2: Annotation Weight Calculation
Semantic Annotation Weight
Step 4: Annotation with Weight
Documents with Semantic Annotations

Figure 4.4 – Proposed Semantic Annotation Algorithm

4.3.5 Error Analysis

One main type of error was identified based on the testing results. Documents that have unbalanced match with the semantic concept index (have many terms from the semantically-related concepts but fewer or no terms from the original concept) were unfairly given high annotation weights. Because most of the terms in the semantic concept index come from the semantically-related concepts, they collectively have a greater impact on the annotation weight despite being penalized in the semantic similarity assessment. For example, the document shown in Figure 4.5 provides general guidance on cumulative impact analysis for NEPA, and is not describing any environmental screening tool/method. However, it was mistakenly annotated as one of the top 10 documents relevant to the concept "environmental screening process". This is because it contains many terms (highlighted in red) from the semantically-related concepts "impact analysis process" and "environmental resources" but no term from the original concept. In future work, the SA method could be improved by penalizing documents that have unbalanced match with the semantic concept index and/or by using different semantic similarity measures to optimize the impacts of different terms when estimating the annotation weights.



Consideration Of Cumulative Impacts In EPA Review of NEPA Documents

The combined, incremental effects of human activity, referred to as cumulative impacts, pose a serious threat to the environment. While they may be insignificant by themselves, cumulative impacts accumulate over time, from one or more sources, and can result in the degradation of important resources. Because federal projects cause or are affected by cumulative impacts, this type of impact must be assessed in documents prepared under the National Environmental Policy Act (NEPA).The guidance focuses on specific issues that are critical in EPA's review of NEPA documents under Section 309 of the Clean Air Act. While there is no "cookbook" method of assessing cumulative impacts, the guidance offers information on what issues to look for in the analysis, what practical considerations should be kept in mind when reviewing the analysis, and what should be said in EPA comments concerning the adequacy of the analysis……

For example, important functions to focus on could include the wetlands' role as a nursery for recreationally and/or commercially valuable aquatic species; its ability to minimize downstream flooding; and its ability to improve water quality……

In addition, assessments of impacts to biological resources generally have been limited to selected game species, federally or state listed threatened and endangered species, and wetlands habitats. These approaches are too limited and should be expanded to consider other valuable resources which could be affected, while also considering a broader array of potential effects.

As an example, federal assessment and mitigation for the loss of wetlands often focus primarily on the acreage affected rather than the function of the wetland within the broader ecosystem. In such a case, the impact to the wetland might not be deemed significant if the wetland had no immediate wildlife values or other notable characteristics……

Figure 4.5 – A Document (Partial) Incorrectly Annotated as one of the Top 10 Relevant Documents

CHAPTER 5: CONTEXT-AWARE INFORMATION RETRIEVAL FOR SUPPORTING ENVIRONMENTAL REVIEW AND PROJECT DECISION MAKING

5.1 State of the Art and Knowledge Gaps

In recent years, a number of important efforts have been conducted for improving document ranking methods for supporting information retrieval (IR) in the construction domain. For example, Soibelman et al. (2007) combined the vector space model (VSM) with document classification information to rank documents related to a project model object, and developed a domain-specific thesaurus to improve the retrieval of construction product information from the internet. Lin and Soibelman (2007; 2009) extended the Boolean model to rank online documents on AEC products based on the similarity between the expanded query vectors and the document vector. Lin et al. (2012) solved the problem of incorrect ranking due to concept density through partitioning technical documents on AEC projects and research into OntoPassages according to domain knowledge, and evaluated the VSM, a probabilistic model, and a language model for ranking OntoPassages. Fan et al. (2015) improved the classical VSM-based document ranking from two perspectives: (1) highlighting the documents containing project-specific information by improving feature weighting based on the project-specific terms and the dependency relations of these terms; and (2) applying machine learning algorithms to optimize the feature weighting based on user feedback. However, all these IR efforts in the construction domain built on keyword-based document ranking methods, which provide limited capabilities for incorporating contextual information into the retrieval process.

Outside of the construction domain, a number of important IR research efforts have been conducted to develop semantic-based document ranking methods. For example, Turney and Pantel

(2010) summarized three different approaches to integrating semantics with the VSM – term-document matrix, word-context matrix, and pair-pattern matrix – and discussed their application for improving IR. Bikakis et al. (2010) proposed an ontology-based IR framework that utilizes a flexible combination of keyword-based and semantic-based document ranking methods, where the semantic-based method is based on the semantic similarity between the target concept(s) and the documents. Fernandez et al. (2011) semantically enhanced the IR using an ontology-based approach, where both the document and the query were represented as vectors of semantic concepts and document ranking was conducted based on the similarity between two concept vectors. Bouramoul et al. (2012) improved the document ranking of current search engines (Google, Bing, and Yahoo) through re-ranking the top retrieved results based on the similarity between the expanded document vector and the query vector, where both vectors were expanded with WordNet concepts linked by semantic relations. AlMasri et al. (2014) tackled the term mismatch problem for document ranking through modifying documents according to a given query and semantic relations between terms, and adapted a number of language models to expand a document by the query terms that have semantically-related document terms but do not appear in the document. Hahm et al. (2015) proposed a semantic-based document ranking approach that incorporates relationships among terms in the relevance assessment process based on a domain ontology, which represents the semantics of a document through a document semantic network and considers both user interests and searching intent through relation-based weighting.

Despite the importance of the above-mentioned research efforts, there still exist many challenges in developing document ranking methods that can efficiently retrieve relevant information for transportation project decision making. In this regard, three major limitations in existing IR research efforts have been identified. First, most of the existing document ranking efforts are limited in their formal context representation – they lack an explicit, domain-specific representation of the concept of context, and can only capture limited semantic information by their document ranking methods. For example, Soibelman et al. (2007) only incorporated document classification information into VSM-based document ranking; Fan et al. (2015) only considered project-specific terms and dependency relations in their relevance evaluation; and AlMasri et al. (2014) only represented hierarchical relations or specific-generic relations between terms in their document ranking method. Second, existing semantic-based document ranking efforts considered limited semantic and contextual information when conducting semantic relevance assessment. For example, Bikakis et al. (2010) only considered the target concept when assessing semantic relevance; while Fernandez et al. (2011) considered both the target concept and the related contextual concepts but treated all the contextual concepts equally without considering their semantic differences. Third, most of the existing semantic-based document ranking efforts build on either the VSM or the statistical language model (SLM), and have not compared the retrieval performance of the two models.

5.2 Proposed Context-Aware Information Retrieval Method

To address the aforementioned gaps, this research proposes a new context-based semantic relevance assessment method that considers the semantic and contextual information of both the target concept and its semantically-related concepts, while taking their semantic relatedness into account through semantic similarity measures. This allows for a deep, context-aware, and semantically-sensitive document representation that better supports document ranking. The proposed method represents the documents with document context concepts in the TPER semantic model and estimates the semantic relevance based on a semantically-extended set of concepts and their relative semantic relatedness to the target concept. The TPER semantic model (see Chapter

4) is a model for representing and reasoning about information and IR in the TPER domain. The document context concepts represent the context dimensions of the document, which describe a collection of relevant conditions and settings that make the semantics of the document unique and comprehensible to that condition, including the project context, the functional process context, and the resource context. To further evaluate which model – the VSM or the SLM – works better for context-enhanced semantic document ranking in the TPER domain, this research further integrates the proposed semantic relevance assessment method into both models. Both, the context-enhanced VSM-based and the SLM-based semantic document ranking methods, were compared with each other and with the original keyword-based methods.

5.2.1 Context-Based Relevance Assessment

The proposed context-based relevance assessment method enhances context-awareness of relevance ranking through an enriched representation of concepts and a deeper and semantically-sensitive estimation of semantic relevance. The proposed method includes three primary elements: semantic concept indexing, semantic relevance estimation, and semantic document representation.

5.2.1.1 Semantic Concept Representation and Indexing

This research proposes a context-aware and deep semantic concept indexing approach to enriching the semantic representation of concepts for enhanced recognition of document relevance. First, the proposed approach improves the representation of context by using a domain-specific context model (i.e., the TPER semantic model). Second, the proposed approach achieves deeper semantic representation by taking the concept terms (i.e., the most common text descriptions of the concept) of both the original concept and its semantically-related concepts (i.e., direct and indirect subconcepts, as well as non-hierarchically-related concepts) into account. For each context concept in the TPER semantic model, a semantic concept index is used to represent its context terms. These terms include the concept terms of the original concept and the concept terms of its semantically-related concepts. The semantic concept index (CI_i) of a concept c_i is represented in Eq. (5.1) (as per Chapter 4), where t_j is an original concept term of c_i, m is the total number of original concept terms of c_i, c_l is a concept that is semantically-related to c_i, t_{lk} is a concept term of c_l, p is the total number of semantically-related concepts to c_i, and q is the total number of concept terms of c_l.

$$CI_i = \{\bigcup_{j=1}^{m} t_j \, , \bigcup_{l=1}^{p} \bigcup_{k=1}^{q} t_{lk}\} \tag{5.1}$$

Figure 5.1 shows an example concept c_i (i.e., “environmental mitigation process”), its semantically-related concepts (including “environmental resource”, “ impact analysis”, “impact avoidance process”, “impact minimization process”, “environmental restore process”, “impact reduction process”, and “environmental compensation process”), the resulting semantic concept index of c_i, and a partial view of the TPER semantic model.

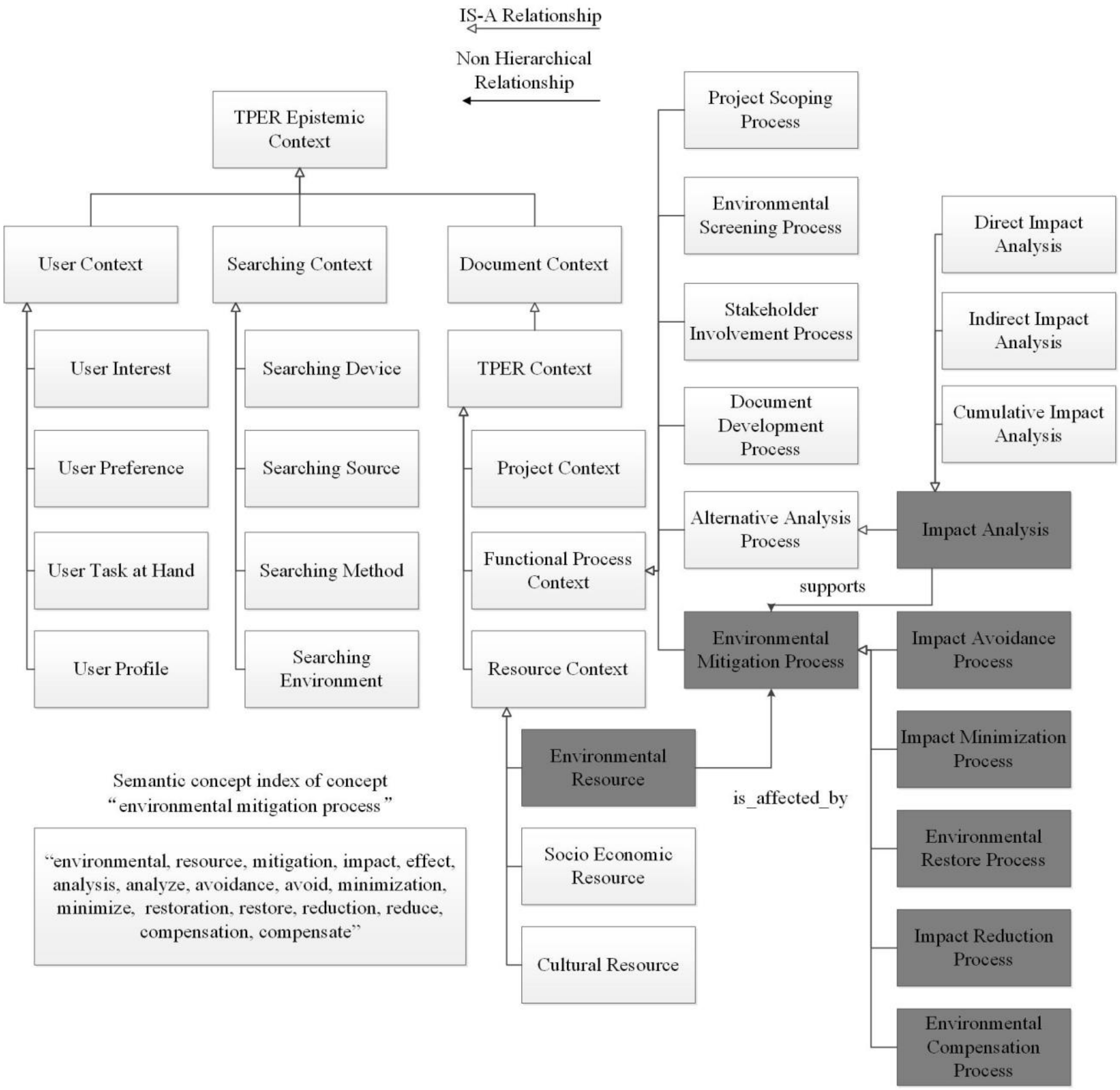

IS-A Relationship
Non Hierarchical Relationship
TPER Epistemic Context
User Context
Searching Context
Document Context
User Interest
User Preference
User Task at Hand
User Profile
Searching Device
Searching Source
Searching Method
Searching Environment
TPER Context
Project Context
Functional Process Context
Resource Context
Project Scoping Process
Environmental Screening Process
Stakeholder Involvement Process
Document Development Process
Alternative Analysis Process
Environmental Mitigation Process
Direct Impact Analysis
Indirect Impact Analysis
Cumulative Impact Analysis
Impact Analysis
supports
Impact Avoidance Process
Impact Minimization Process
Environmental Restore Process
Impact Reduction Process
Environmental Compensation Process
Environmental Resource
is_affected_by
Socio Economic Resource
Cultural Resource
Semantic concept index of concept "environmental mitigation process"
"environmental, resource, mitigation, impact, effect, analysis, analyze, avoidance, avoid, minimization, minimize, restoration, restore, reduction, reduce, compensation, compensate"

Figure 5.1 – The Semantically-Related Concepts and Semantic Concept Index of the Concept "Environmental Mitigation Process"

5.2.1.2 Semantic Relevance Estimation

This research proposes a deep and semantically-sensitive relevance estimation approach. First, the proposed approach achieves a deeper level of semantic relevance assessment by representing the original query through a semantically-extended set of concepts [the target concept (in a query) and its semantically-related concepts]. Second, the proposed approach is semantically-sensitive by considering the relative semantic relatedness of these semantically-related concepts to differentiate their level of relevance to the original query. For each document, its semantic relevance to a context concept is estimated based on two factors: (1) the semantic relatedness between the target concept and its semantically-related concepts, and (2) the occurrence of the context terms in the document. The research proposes a concept relatedness vector [$\vec{R}(CI_i)$] and a concept weight vector [$\overrightarrow{W}(c_i|d_n)$] to represent these two factors, respectively.

Concept relatedness is represented by term relatedness factors (TRFs), which measure the degrees of relatedness between semantically-related concept terms and an original concept. A TRF is a measure of semantic relevance of a term to a concept. For original concept terms, the TRF value equals to 1. For the concept terms of semantically-related concepts, the TRF value is measured in terms of semantic similarity between the original concept and its semantically-related concept. The concept relevance vector, $\vec{R}(CI_i)$, of a concept c_i is expressed in Eq. (5.2), where c_i is the original concept, t_j is an original concept term of c_i, $TRF(t_j)$ is the TRF of t_j and equals to 1, c_l is a concept that is semantically-related to c_i, t_{lk} is a concept term of c_l, $TRF(t_{lk})$ is the TRF of t_{lk} that is calculated as the normalized semantic similarity (SS) between c_i and c_l and is measured using Al-Mubaid and Nguyen (2006) SS measure, *m* is the total number of original concept terms of c_i, p is the total number of semantically-related concepts to c_i, and q is the total number of concept terms of c_l. Al-Mubaid and Nguyen (2006) SS measure estimates the SS between two concepts based on the shortest path between two concepts and the common specificity (CSpec) of the two concepts, which indicates how much common information two concepts share.

$$\vec{R}(CI_i) = \{\cup_{j=1}^{m} TRF(t_j), \cup_{l=1}^{p} \cup_{k=1}^{q} TRF(t_{lk})\} \tag{5.2}$$

The concept weight vector represents the discriminating power of the context terms in the semantic concept index, which is measured by the frequency-inverse document frequency (TF-IDF) term weight. For a document d_n and a concept c_i, the concept weight vector $\vec{W}(c_i|d_n)$ is defined using Eq. (5.3), where W_{t_j} is the TF-IDF weight of concept term t_j of c_i, *m* is the total number of original concept terms of c_i, c_l is a concept that is semantically-related to c_i, $W_{t_{lk}}$ is the TF-IDF weight of the concept term t_{lk} of c_l, p is the total number of semantically-related concepts to c_i, and q is the total number of concept terms of c_l.

For document d_n and concept c_i, the semantic relevance $S_{c_i}(d_n)$ is defined using Eq. (5.4) (Nesic et al. 2010), where $\vec{W}(c_i|d_n)$ is the concept weight vector of c_i, and $\vec{R}(CI_i)$ is the concept relevance vector of c_i.

$$\vec{W}(c_i|d_n) = \{\bigcup_{j=1}^{m} W_{t_j}, \bigcup_{l=1}^{p} \bigcup_{k=1}^{q} W_{t_{lk}}\} \tag{5.3}$$

$$S_{c_i}(d_n) = \vec{W}(c_i|d_n) * \vec{R}(CI_i) \tag{5.4}$$

5.2.1.3 Sematic Document Representation

Because of the proposed approaches to semantic concept representation and relevance estimation, the proposed method is able to represent a document in a deep, context-aware, and semantically-sensitive manner. A document is represented in terms of document context concepts and their semantic relevance to the document. For each document in the collection, its semantic relevance to every document context concept in the TPER semantic model is estimated – in a deep and semantically-sensitive way (as described in Section 5.2.1.2) – to create the context representation of the document. For a document d_n, its context representation is defined as a document concept vector $\vec{C}(d_n)$, and is shown in Eq. (5.5), where $S_{c_i}(d_n)$ is the semantic relevance of concept c_i to document d_n, and H is the total number of document context concepts in the TPER semantic model.

$$\vec{C}(d_n) = \{\cup_{i=1}^{H} S_{c_i}(d_n)\} \tag{5.5}$$

5.2.2 Integrating the Proposed Relevance Assessment Method into Document Ranking Models

Both, the original VSM and the original SLM build on keyword-based document representation and query processing techniques, and rely on term relevance to conduct document ranking. To enable context-based semantic document ranking, the proposed semantic relevance assessment method was integrated into these document ranking methods, in both semantic query processing and semantic relevance ranking. These two document ranking methods were selected for further study, because they are the most widely used. The VSM and the SLM each offers different advantages in different situations. Previous studies (Lin et al. 2012; Bennett et al. 2008; Raghayan and Iyer 2007) indicated that there is no single model that outperforms the other in all applications. For example, Raghavan and Iyer (2007) found that the VSM had better performance when retrieving relevant advertisement for sponsored search; while Lin et al. (2012) found that the SLM was better at retrieving passages of technical documents for AEC projects and research. Because the performances of the two models could vary from domain to domain and application to application, it is necessary to compare the performances of the two models in facilitating context-enhanced semantic document ranking in the TPER domain.

5.2.2.1 Semantic Query Processing

Semantic query processing (SQP) provides the context representation of a user's query by extracting context concepts from the query and transforming it into a semantic query. A semantic query consists of document context concepts in the TPER semantic model whose concept term(s) appear in the query. For the query, the concept terms in the concept index of every concept are searched (using term-based matching) to check if they appear in the query. If a concept term appears in the query, its corresponding concept is added into the corresponding semantic query. For example, as per Figure 5.2, for the query "how to assess corridor alignments for effect on traffic congestion", the semantic query is "alternative analysis process, highway project, impact analysis, mobility".

For each semantic query Q_t, a query concept vector $\vec{C}(Q_t)$ is defined using Eq. (5.6), where $I_{c_i}(Q_t)$ is the concept indicator that represents how important the user values a concept c_i, and H is the total number of document context concepts in the TPER semantic model. For the testing queries, the concept indicator of each concept in the corresponding semantic query was set to 1, based on the assumption that the user gives equal importance to these concepts.

$$\vec{C}(Q_t) = \{\cup_{i=1}^{H} I_{c_i}(Q_t)\} \tag{5.6}$$

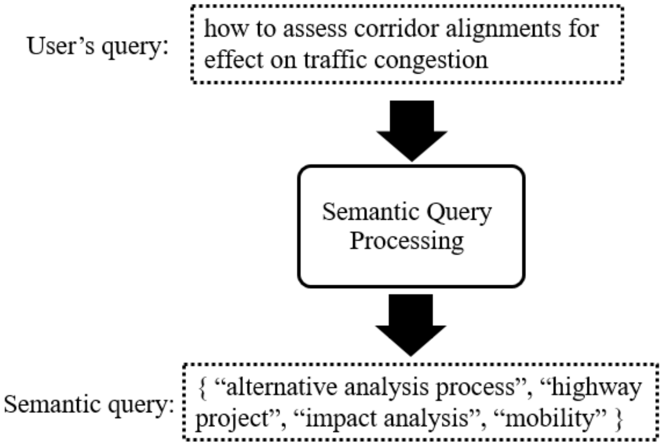

User's query:
how to assess corridor alignments for effect on traffic congestion
Semantic Query Processing
Semantic query:
{ "alternative analysis process", "highway project", "impact analysis", "mobility" }

Figure 5.2 – An Example of a User's Query and its Corresponding Semantic Query

5.2.2.2 Semantic Relevance Ranking

Integrating the proposed context-based relevance assessment method in the VSM and the SLM requires different ways to estimate semantic relevance between a document and a query based on their context representations and different ways to integrate semantic relevance with their original term relevance, because each model has a different notion about what does relevance mean. Based on these differences, two context-enhanced semantic document ranking methods were proposed: VSM-based method and SLM-based method.

5.2.2.2.1 Context-Enhanced Vector Space Model-Based Method

For the proposed context-enhanced VSM-based method, (1) term relevance is measured by term similarity, which is the similarity between a query and a document at the term level; and (2) semantic relevance is measured by context similarity, which is the similarity between the query's corresponding semantic query and the context representation of the document.

Term similarity is a measure of the cosine similarity between the document term vector and the query term vector. To measure term similarity, a document d_n and a query Q_t should, thus, be represented as a document term vector and a query term vector, respectively. For each document d_n, its document term vector $\vec{T}(d_n)$ is defined using Eq. (5.7) (Roelleke 2013), where $W_{t_g}(d_n)$ is the TF-IDF weight of a unique term t_g in document d_n, and x_d is the total number of unique terms in the document collection.

$$\vec{T}(d_n) = \{\cup_{g=1}^{x_d} W_{t_g}\,(d_n)\} \qquad (5.7)$$

For query Q_t, its query term vector $\vec{T}(Q_t)$ is defined using Eq. (5.8) (Roelleke 2013), where $W_{t_g}(Q_t)$ is the TF-IDF weight of a unique term t_g in query Q_t, and x_d is the total number of unique terms in the document collection.

$$\vec{T}(Q_t) = \{\cup_{g=1}^{x_d} W_{t_g}(Q_t)\} \tag{5.8}$$

Accordingly, the term similarity $sim^t(Q_t, d_n)$ between document d_n and query Q_t is defined using Eq. (5.9) (Aggarwal and Zhai 2012), where $\vec{T}(Q_t)$ is the query term vector for Q_t, $\vec{T}(d_n)$ is the document term vector for d_n, $||\vec{T}(Q_t)||$ is the length of the query term vector, and $||\vec{T}(d_n)||$ is the length of the document term vector.

$$sim^t(Q_t, d_n) = \frac{\vec{T}(Q_t) * \vec{T}(d_n)}{||\vec{T}(Q_t)|| * ||\vec{T}(d_n)||} \tag{5.9}$$

To better incorporate contextual information in ranking documents, the use of context similarity is proposed in this research, in order to measure the relevance of a document to a query based on the similarity between their context representations. Documents differ in terms of their contextual information (i.e., relevant document context concepts and semantic relevance), where a document with a higher context similarity to a query indicates a higher relevance to that query. The proposed context similarity is a measure of the cosine similarity between the document concept vector and the query concept vector, where all concepts are context concepts from the TPER semantic model. The proposed similarity equation [Eq. (5.10)] defines the context similarity $sim^c(Q_t, d_n)$ between document d_n and query Q_t, where $\vec{C}(Q_t)$ is the query concept vector for Q_t, $\vec{C}(d_n)$ is the document concept vector for d_n, $||\vec{C}(Q_t)||$ is the length of the query concept vector and $\|\vec{C}(d_n)||$ is the length of the document concept vector.

$$sim^c(Q_t, d_n) = \frac{\vec{C}(Q_t) * \vec{C}(d_n)}{||\vec{C}(Q_t)|| * ||\vec{C}(d_n)||} \tag{5.10}$$

Accordingly, document relevance to a query is defined in terms of, both, context similarity and term similarity, where a factor (0 to 1) is used to control the contributions of context similarity and term similarity to document relevance. The proposed relevance equation [Eq. (5.11)] defines the relevance $R(d_n|Q_t)$of document d_n to query Q_t, where $sim^c(Q_t, d_n)$ is the context similarity between document d_n and query Q_t, $sim^t(Q_t, d_n)$ is the term similarity between document d_n and query Q_t, and α is the contribution factor that controls the contributions of context similarity and term similarity to document relevance. In order to find the optimized contribution factor, a range of values (α from 0 to 1 with intervals of 0.1) were tested, and $\alpha = 0.6$ was used for the experiments conducted in this research.

$$R(d_n|Q_t) = \alpha\, sim^c(Q_t, d_n) + (1 - \alpha) sim^t(Q_t, d_n) \tag{5.11}$$

5.2.2.2.2 Context-Enhanced Statistical Language Model-Based Method

In the proposed context-enhanced SLM-based method, (1) term relevance is measured by term probability, which is the probability that a document is relevant to a query at the term level; and (2) semantic relevance is measured by context probability, which is the probability that a document is relevant to a query at the context level.

Term probability is the conditional probability that a document is relevant given a certain query. Given a user's query Q_t and a document d_n, the term probability $P(d_n|Q_t)$ was derived using Bayes rule in Eq. (5.12) (Zhai 2008), where $P(Q_t|d_n)$ is the posterior probability that a user who likes to retrieve document d_n would use query Q_t, $P(d_n)$ is the document prior probability that document d_n is relevant to any query (i.e., it is a document-specific probability that is query-independent), and $P(Q_t)$ is the probability that a user uses query Q_t. Assuming the user has no preference towards any document and $P(Q_t)$ is a constant for a given query Q_t, the term probability $P(d_n|Q_t)$ was treated as equal to the posterior probability $P(Q_t|d_n)$ when measuring the term relevance of documents.

$$P(d_n|Q_t) = \frac{P(Q_t|d_n)*P(d_n)}{P(Q_t)} \tag{5.12}$$

The posterior probability $P(Q_t|d_n)$ is defined using Eq. (5.13) (Manning et al. 2009), where t_g is a term that appears in query Q_t, $p(t_g|\theta_n)$ is the probability of generating term t_g according to document language model θ_n of document d_n, and x_q is the total number of terms in query Q_t. The document language model θ_n is a probability distribution of terms given document d_n (Buttcher et al. 2010). As the most successful and popular language model, the unigram multinomial language model was adopted for θ_n.

$$P(Q_t|d_n) = \prod_g^{x_q} p(t_g|\theta_n) \tag{5.13}$$

The probability $p(t_g|\theta_n)$ is defined using Eq. (5.14) (Singhal 2001), where $tf(t_g, d_n)$ is the frequency of term t_g in document d_n, $||d_n||$ is the length of document d_n, $\mathrm{p}(t_g|\theta_b)$ is the probability of generating term t_g according to the background language model θ_b, and λ is a smoothing factor that controls the contribution of the background language model. The background language model θ_b is a probability distribution of terms given the entire document collection. In order to find the optimized smoothing factor, a range of values (λ from 0 to 1 with intervals of 0.1) were tested, and $\lambda = 0.3$ was used for the experiments conducted in this research. The probability $\mathrm{p}(t_g|\theta_b)$ is defined using Eq. (5.15) (Zhai 2008), where $tf(t_g, d_c)$ is the frequency of term t_g in document collection d_c, and $||d_c||$ is the total length of the document collection d_c.

$$p(t_g|\theta_n) = (1-\lambda)\frac{tf(t_g,d_n)}{||d_n||} + \lambda\, \mathrm{p}(t_g|\theta_b) \tag{5.14}$$

$$\mathrm{p}\left(t_g|\theta_b\right) = \frac{tf(t_g,d_c)}{||d_c||} \tag{5.15}$$

To better incorporate contextual information in ranking documents, the use of context probability is proposed in this research, for the SLM-based model, in order to measure the relevance of a document to a query based on the likelihood that the document is relevant to a query on the contextual level. In this case, a document with a higher context probability to a query indicates a higher relevance to that query. The proposed context probability is a probability measure based on context similarity; it measures the likelihood that a document is relevant to a query based on the context similarity between that document and that query relative to the aggregated context similarities of all documents to that query. The proposed context probability equation is defined in Eq. (5.16), where $P(d_n|Q_t)^c$ is the context probability that document d_n is relevant to query Q_t, $sim^c(Q_t, d_n)$ is the context similarity between query Q_t and document d_n, and N is the total number of documents in the collection.

$$P(d_n|Q_t)^c = \frac{sim^c(Q_t,d_n)}{\sum_{n=1}^{N} sim^c(Q_t,d_n)} \tag{5.16}$$

Accordingly, document relevance to a query is defined in terms of, both, context probability and term probability. The proposed probability-based relevance equation [Eq. (5.17)] defines the relevance $R(d_n|Q_t)$ of document d_n to query Q_t, where $P(d_n|Q_t)$ is the term probability, and $P(d_n|Q_t)^c$ is the context probability.

$$R(d_n|Q_t) = P(d_n|Q_t) * P(d_n|Q_t)^c \tag{5.17}$$

5.2.3 Experimental Setup

A set of experiments were conducted to: (1) evaluate the effectiveness of the proposed context-based relevance assessment method, and (2) evaluate the context-enhanced semantic document ranking methods and compare their IR performance. The following subsections explain the data preparation, data preprocessing, and evaluation efforts.

5.2.3.1 Data Preparation

A collection of textual documents in the TPER domain was first created. To create the document collection, the domains of the following categories of websites were crawled: (1) websites on environmental review process guidelines, including the FHWA Environmental Review Toolkit **(**www.environment.fhwa.dot.gov) and the Center for Environmental Excellence by American Association of State Highway and Transportation Officials (AASHTO) (http://environment.transportation.org); (2) websites of environmental review process stakeholders, including the IDOT (http://www.idot.illinois.gov), the Environmental Protection Agency (EPA) (http://www.epa.gov), the Illinois Department of Natural Resources (IDNR) (http://www.dnr.illinois.gov), and the Illinois Historic Preservation Agency (IHPA) (http://www.illinois.gov/ihpa); and (3) websites of large-scale transportation projects, such as the Illiana Corridor (http://www.illianacorridor.org) and the Eisenhower Expressway (http://eisenhowerexpressway.com/). The homepages of the above-described websites were compiled as a list of seed pages. Starting from every seed page, a web crawler was utilized to examine every web page under the domain; extract its URL, headings, and the body text; and save them in a .txt format local file. The final document collection contained 5,436 documents.

5.2.3.2 Data Preprocessing

Three data preprocessing techniques were utilized: tokenization, stopword removal, and lemmatization. Tokenization breaks the text into meaningful units (i.e., tokens). For the experiments conducted in this research, a token was defined as a single word. Stopword removal removes words that have high frequency but have low power in discriminating documents that match a user need (i.e., stopwords). Removing stopwords reduces the number of features, reveals discriminative words, and in turn improves IR performance. Lemmatization transforms a word into its base or dictionary form (i.e., lemma). Lemmatization can reduce the number of features by grouping the words with the same lemma, and can in turn be effective in enhancing IR performance.

5.2.3.3 Testing Queries

To evaluate the performance of different document ranking methods, a set of testing queries were developed based on interactions with a set of 31 transportation project practitioners during one-to-one interviews, including practitioners from the following agencies: IDOT districts, MPOs, resource agencies, IDOT central office, and FHWA (ICT 2014). All the interviewed experts are practitioners involved in conducting, supervising, and/or coordinating planning and/or environmental studies (e.g., planning director, environmental study supervisor), and 30 out of the 31 experts have over 10 years of relevant working experience (ICT 2014). These experts were targeted as they are more familiar with the roles and responsibilities of different agencies in the transportation planning process and/or NEPA process and can provide better feedback on the identified testing queries. A total of 18 testing queries were developed. As shown Table 5.1, based on the length of the query, the 18 testing queries were further classified into two groups: short and long queries. A short query contains fewer than five terms, whereas a long query includes five or more terms. The testing queries represent the basic information needs from transportation professionals during the environmental review process, and cover the important information-seeking tasks of the process (ICT 2014), such as estimating potential environmental impact, developing mitigation measures, and preparing environmental permits.

Table 5.1 – The Testing Queries

Query number	Query description	Query classification
1	Mitigation measures	Short
2	Environmental screening method	Short
3	Estimate environmental impact	Short
4	GIS data	Short
5	Highway corridor project	Short
6	Wetland section 404 permit	Short
7	Mitigation measures for wetland resource	Long
8	Environmental screening method for highway project	Long
9	Estimate environmental impact on air quality	Long
10	GIS data for historic resource	Long
11	Highway corridor project in Illinois	Long
12	Wetland section 404 permit for highway project	Long
13	Mitigation measures for wetland in Illinois	Long
14	Environmental screening method for highway project in Florida	Long
15	Estimate environmental impact on air quality for highway project	Long
16	GIS data for historic resource in Illinois	Long
17	Highway corridor project in Illinois with NEPA study	Long
18	Wetland section 404 permit for highway project in Illinois	Long

5.2.4 Evaluation

Given the specificity of the information needs and the size of the document collection, manually judging the relevance of each document is a time-consuming process. To improve the efficiency of relevance assessment, "pooling" – a non-exhaustive assessment method – is commonly adopted. Using pooling, "relevance is assessed over a subset of the collection that is formed from the top k documents returned by a number of different information retrieval systems" (Manning et al. 2009). In the experiments conducted for this research, for each query, the top 50 documents retrieved using the two different semantic document ranking methods (context-enhanced VSM-based and SLM-based methods) and their provenance methods (original VSM and SLM) were pooled together and manually assessed by three judges (the author and two other researchers). Each judge independently assessed each document in the pool to determine whether it is relevant to the query or not. For each document, the relevance judgement was based on the agreement between judges. Two main methods were used for discrepancy resolution: (1) If a majority (i.e., at least two) of the judges achieved agreement, then the agreed-on judgement was used; and (2) If a majority of the judges did not achieve agreement, then a discussion was conducted until a majority agreement was achieved.

For a query Q_t, each document in the collection was ranked based on the relevance score $R(d_n|Q_t)$. The performance of the context-enhanced VSM-based and SLM-based methods were evaluated using MP at the top k retrieved documents and MAP. For a query Q_t, precision at rank *k* was calculated based on Eq. (5.18), where RT_k is the number of relevant documents retrieved at rank *k*, and RE_k is the total number of documents retrieved at rank *k*. MP for a set of queries is the arithmetic mean of the precision values of the queries. MP at rank k was calculated as per Eq. (5.19), where $P(k)_t$ is the precision of query Q_t at rank *k*, and *B* is the total number of queries.

MAP was calculated based on precision and average precision (AP). AP for query Q_t was calculated as per Eq. (5.20), where k is the rank of a document based on the relevance score $R(d_n|Q_t)$ for query Q_t, $P(k)_t$ is the precision of query Q_t at rank *k*, *rel(k)* is an indicator function that equals to 1 if the retrieved document at rank k is relevant and 0 otherwise, RL is the total number of relevant documents, and *RT* is the total number of retrieved documents. MAP for a set of queries is the arithmetic mean of the AP values of the queries. Accordingly, MAP was calculated as per Eq. (5.21), where *AP(Q_t)* is the average precision of query Q_t and *B* is the total number of queries. For the experiments conducted in this research, MP values at the top 10, 20, 30, 40, and 50 retrieved documents and MAP were calculated.

$$P(k)_t = \frac{RT_k}{RE_k} \tag{5.18}$$

$$MP(k) = \frac{\sum_{t=1}^{B} P(k)_t}{B} \tag{5.19}$$

$$AP(Q_t) = \frac{\sum_{k=1}^{RT} P(k)_t * rel(k)}{RL} \tag{5.20}$$

$$MAP = \frac{\sum_{t=1}^{B} AP(Q_t)}{B} \tag{5.21}$$

5.3 Experimental Results and Analysis

The proposed context-enhanced semantic document ranking methods were tested in retrieving webpages that are relevant to TPER. The methods were tested on a testing data set of 5,436 Web pages (as discussed in Section 5.2.3.1). The evaluation focused on testing the two proposed context-enhanced document ranking methods: the VSM-based method (with context similarity) and the SLM-based method (with context probability). First, each context-enhanced method was compared to its provenance method: the original VSM-based method (keyword-based) and the original SLM-based method (keyword-based), respectively. Second, both context-enhanced methods were compared with each other.

5.3.1 Performance of Vector Space Model-Based Methods

To conduct the first comparative evaluation, document ranking was conducted in two different ways: (1) using the original VSM-based method, and (2) using the proposed context-enhanced VSM-based method. The performance results of the two methods are summarized in Table 5.2.

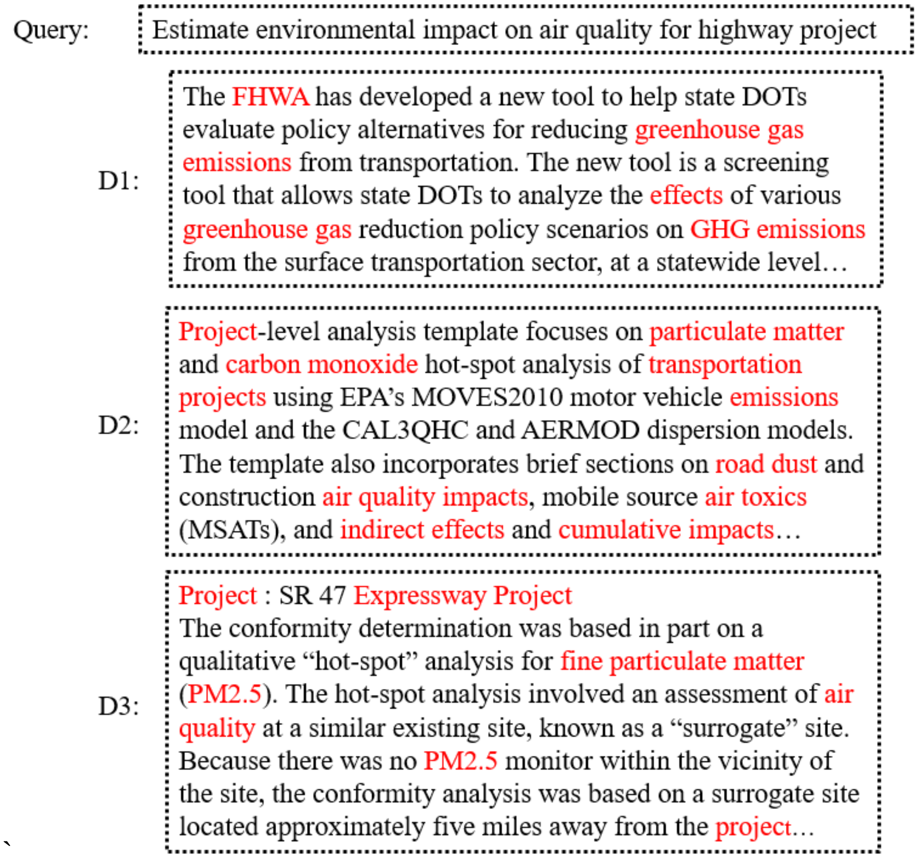
Table 5.2 – The Performance of the Original and the Context-Enhanced Vector Space Model (VSM)-Based Methods

Query group	Original VSM-based method						Context-enhanced VSM-based method					
	MP[a] at top 10	MP[a] at top 20	MP[a] at top 30	MP[a] at top 40	MP[a] at top 50	MAP[b]	MP[a] at top 10	MP[a] at top 20	MP[a] at top 30	MP[a] at top 40	MP[a] at top 50	MAP[b]
Short queries	80%	71%	68%	68%	65%	40%	82%	80%	81%	80%	79%	54%
Long queries	63%	55%	52%	48%	46%	32%	78%	65%	63%	60%	57%	45%
Performance difference[c]	21%	22%	23%	29%	29%	21%	4%	19%	22%	25%	28%	16%
Overall	69%	61%	57%	55%	53%	35%	79%	70%	68%	66%	65%	48%

[a] MP: mean precision
[b] MAP: mean average precision
[c] Performance difference (%) = (absolute difference/original performance) x 100

As shown in Table 5.2, using the original VSM-based method, the overall MAP was 35%, and the MPs at the top 10, 20, 30, 40, and 50 retrieved documents were 69%, 61%, 57%, 55%, and 53%, respectively. For the two query groups, the performance dropped for the long queries by 21% for MAP, and by 21%, 22%, 23%, 29%, and 29% for MPs at the top 10, 20, 30, 40, and 50 retrieved documents, respectively. This decrease in performance could be due to the fact that as the query becomes more specific (i.e., the length of the query increases), the number of relevant documents in the collection becomes smaller and thus relevant documents become harder to retrieve. For example, the average number of relevant documents in the pool for short queries is 72, which drops to 41 for long queries.

On the other hand, as shown in Table 5.2, using the context-enhanced VSM-based, the overall MAP was improved to 48%, and the MPs at the top 10, 20, 30, 40, and 50 retrieved documents were improved to 79%, 70%, 68%, 66%, and 65%, respectively. Because AP integrates both precision and recall, AP values were analyzed to evaluate whether the performance improvement after adopting the context-enhanced VSM-based method is statistically significant. The paired student’s t-test was used to evaluate the improvement in AP for all 18 testing queries. The results of the t-test were interpreted based on the probability value (p-value). If the p-value is less than 0.05, then the difference is statistically significant. The p-value for the APs of the 18 testing queries was 0.006, which indicates that the context-enhanced VSM-based method significantly improves AP in comparison to the status-quo method (the original VSM-based method). These results show that the use of context similarity as a measure of document relevance to queries is effective in improving IR performance. This is because the context similarity can capture semantically-related terms that are otherwise ignored by the original VSM-based method. For example, Figure 5.3 shows the top 3 retrieved documents (partial) from the query “estimate environmental impact on air quality for highway project” using the context-enhanced VSM-based method, and the terms highlighted in red color are the terms that contribute to the context similarity between each document and the query. Many semantically-meaningful terms that do not appear in the query, such as “greenhouse gas”, “particulate matter”, and “carbon monoxide” will not be captured using the original VSM-based method.



Query: Estimate environmental impact on air quality for highway project

D1: The FHWA has developed a new tool to help state DOTs evaluate policy alternatives for reducing greenhouse gas emissions from transportation. The new tool is a screening tool that allows state DOTs to analyze the effects of various greenhouse gas reduction policy scenarios on GHG emissions from the surface transportation sector, at a statewide level…

D2: Project-level analysis template focuses on particulate matter and carbon monoxide hot-spot analysis of transportation projects using EPA's MOVES2010 motor vehicle emissions model and the CAL3QHC and AERMOD dispersion models. The template also incorporates brief sections on road dust and construction air quality impacts, mobile source air toxics (MSATs), and indirect effects and cumulative impacts…

D3: Project : SR 47 Expressway Project
The conformity determination was based in part on a qualitative "hot-spot" analysis for fine particulate matter (PM2.5). The hot-spot analysis involved an assessment of air quality at a similar existing site, known as a "surrogate" site. Because there was no PM2.5 monitor within the vicinity of the site, the conformity analysis was based on a surrogate site located approximately five miles away from the project…

Figure 5.3 – The Top Ranked Documents (Partial) Retrieved by a Sample Query Using the Proposed Context-Enhanced VSM-Based Method

Compared to the state-of-the-art efforts, the extent of improvement is quite significant (39% improvement for MAP and 15% improvement for MP at the top 10); existing IR efforts (Abbasi and Frommholz et al. 2015; Wang and Akella 2015; Gupta et al. 2014; Han et al. 2014; Babashzadeh et al. 2013) typically show a performance improvement that ranges between 18%-29% for MAP and 10%-19% for MP at the top 10 retrieved documents. The extent of improvement also remains steady across the different MP metrics, ranging from 15% to 23% improvement. Such steady performance is especially important for supporting TPER and project decision making, because transportation practitioners – unlike general (non-specialized) users – tend to search for a large number of relevant documents (e.g., all environmental review studies of similar projects).

For the two query groups, the performance of the context-enhanced method showed a similar dropping trend for long queries, as seen for the original VSM-based method. However, compared with the original method, the extent of performance drop for the context-enhanced method was smaller. The performance of the proposed method dropped for the long queries by 16% for MAP, and by 4%, 19%, 22%, 25%, and 28% for MPs at the top 10, 20, 30, 40, and 50 retrieved documents, respectively. As the query becomes more specific (i.e., the length of the query increases), it is more likely to include context descriptions; and, therefore, the context-enhanced document ranking was able to compensate for the natural drop usually seen when queries become more specific (due to the decrease in the number of relevant documents). Such more robust IR performance when moving from shorter to longer queries is especially important for this domain-specific application, because transportation practitioners tend to have specific information needs that usually involve multiple query terms (and in turn context concepts).

5.3.2 Performance of Statistical Language Model-Based Methods

To conduct the second comparative evaluation, SLM-based document ranking was conducted in two different ways: (1) using the original SLM-based method, and (2) using the proposed context-enhanced SLM-based method. The performance results of the two methods are summarized in Table 5.3.

Table 5.3 – The Performance of the Original and the Context-Enhanced Statistical Language Model (SLM)-Based Methods

Query group	Original SLM-based method						Context-enhanced SLM-based method					
	MP[a] at top 10	MP[a] at top 20	MP[a] at top 30	MP[a] at top 40	MP[a] at top 50	MAP[b]	MP[a] at top 10	MP[a] at top 20	MP[a] at top 30	MP[a] at top 40	MP[a] at top 50	MAP[b]
Short queries	68%	72%	70%	70%	68%	38%	70%	75%	73%	69%	69%	40%
Long queries	56%	51%	50%	48%	45%	28%	60%	54%	52%	50%	48%	31%
Performance difference[c]	18%	29%	29%	31%	34%	27%	14%	28%	28%	27%	30%	22%
Overall	60%	58%	57%	55%	52%	31%	63%	61%	59%	56%	55%	34%

[a] MP: mean precision
[b] MAP: mean average precision
[c] Performance difference (100%) = (absolute difference/original performance) x 100

As shown in Table 5.3, using the original SLM-based method, the overall MAP was 31%, and the MPs at the top 10, 20, 30, 40, and 50 retrieved documents were 60%, 58%, 57%, 55%, and 52%, respectively. For the two query groups, similar to the VSM-based methods, the performance dropped for the long queries; it dropped by 27% for MAP, and by 18%, 29%, 29%, 31%, and 34% for MPs at the top 10, 20, 30, 40, and 50 retrieved documents, respectively.

On the other hand, as shown in Table 5.3, using the context-enhanced SLM-based method, the overall MAP was improved to 34%, and the MPs at the top 10, 20, 30, 40, and 50 retrieved documents were improved to 63%, 61%, 59%, 56%, and 55%, respectively. Similar to the VSM-based methods, AP was analyzed to evaluate whether the performance improvement after adopting the proposed SLM-based method is statistically significant. The paired student's t-test was used to evaluate the improvement in AP for all 18 testing queries. The p-value for the APs of the 18 testing queries was 0.00028, which indicates that the context-enhanced SLM-based method significantly improves AP in comparison to the status-quo method (the original SLM-based method). These results show that the use of context information – here context probability – also improved the IR performance when using an SLM-based method. Conducting a similar comparison to the state-of-the-art (as that in Section 5.3.1), the extent of context-induced improvement for the SLM-based is not as large though: 10% improvement for MAP and 6% improvement for MP at the top 10 for the SLM-based, in comparison to 39% improvement for MAP and 15% improvement for MP at the top 10 for the VSM-based. This could be attributed to the different natures of both types of methods; the VSM is similarity-based, while the SLM is probability-based.

For the two query groups, the performance of the proposed method showed a similar dropping trend for long queries, as seen for the original SLM-based method. However, compared with the original method, the extent of performance drop for the proposed method was smaller. The performance of the proposed method dropped for the long queries by 22% for MAP, and by 14%, 28%, 28%, 27%, and 30% for MPs at the top 10, 20, 30, 40, and 50 retrieved documents, respectively. Similar to the VSM-based method, as the query becomes more complex, the natural performance drop is compensated by the use of context probability.

5.3.3 Comparison of the Context-Enhanced Vector Space Model-Based and Statistical Language Model-Based Methods

To conduct the third comparative evaluation, the two context-enhanced methods were compared. The performance of the context-enhanced VSM-based (with context similarity) and the context-enhanced SLM-based (with context probability) methods are summarized in Table 5.4. As shown in Table 5.4, the context-enhanced VSM-based method outperformed the context-enhanced SLM-based method on every performance metric. To evaluate if the higher performance is statistically significant, the student's t test was used to test the improvements in AP for all 18 testing queries. The p-value for the APs of the 18 testing queries is 0.00014, which indicates that the IR performance of the context-enhanced VSM-based method is significantly better than that of the context-enhanced SLM-based method. The higher performance of the VSM-based method could be attributed to the following two reasons: (1) the context-enhanced VSM-based method introduced a contribution factor that allows for parameter tuning to optimize the contributions of term relevance and semantic relevance while the SLM-based method treated the term relevance and semantic relevance equally; and (2) the performance of the SLM-based method depends largely on the probability estimation based on the document language model, which is very sensitive to noisy data.

Table 5.4 – The Performance of the Proposed Context-Enhanced Vector Space Model (VSM)-Based and Statistical Language Model (SLM)-Based Methods

Document ranking method	MP[a] at top 10	MP[a] at top 20	MP[a] at top 30	MP[a] at top 40	MP[a] at top 50	MAP[b]
Context-enhanced VSM-based method	79%	70%	68%	66%	65%	48%
Context-enhanced SLM-based method	63%	61%	59%	56%	55%	34%

[a] MP: mean precision
[b] MAP: mean average precision

5.3.4 Comparison to the State-of-the-Art Performance

Based on the experimental results, the proposed VSM-based method was selected to better incorporate contextual information in ranking documents. As shown in Table 5.4, the method achieved 48% MAP, 79% MP at the top 10 retrieved documents, and 65% MP at the top 50 retrieved documents. Compared to the performance of the state-of-the-art bibliographic search engines in other domains, the proposed context-enhanced document ranking method showed a relatively high level of performance. A study on the performance of state-of-the-art bibliographic search engines showed that MPs at the top 10 documents for the best-performing systems [i.e., Google Scholar, PubMed, and Social Sciences Citation Index (SSCI)] are between 60% and 80%, and MPs at the top 50 documents are between 32% and 48% (Walters 2011). Compared to these ranges, the proposed context-enhanced document ranking method achieved a high-end performance at the top 10 retrieved documents (79% MP), and an above-range performance (65%

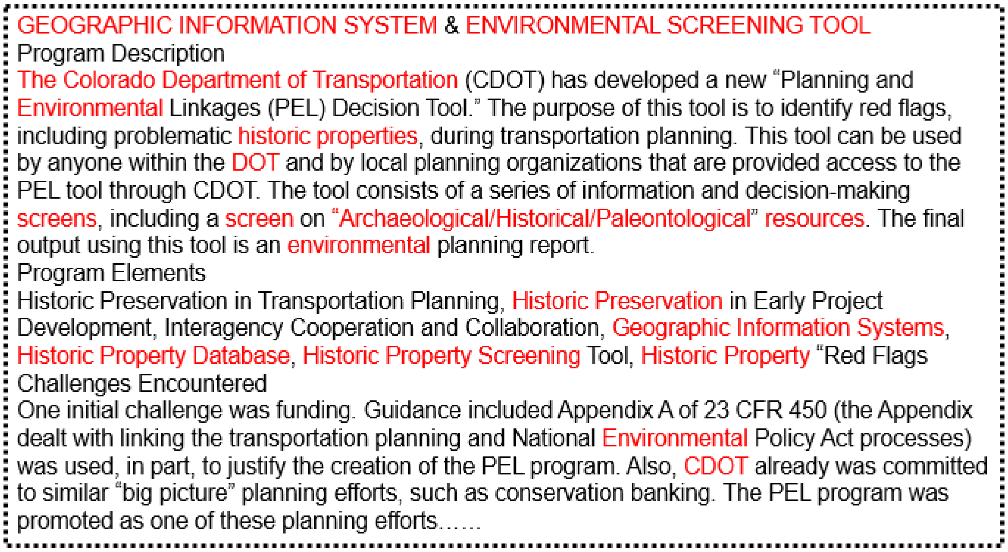
MP) at the top 50 retrieved documents. This indicates that the proposed domain-specific, context-enhanced document ranking method is potentially effective in retrieving information that is more relevant to TPER and decision making.

5.3.5 Scalability

In terms of the scalability, the time efficiency of the proposed context-enhanced document ranking method depends largely on semantic relevance ranking, because context-based relevance assessment can be pre-conducted and the document concept vectors can be stored as metadata for the document collection, which requires updates only when the size of the collection increases or the semantic model changes. For the proposed VSM-based semantic relevance ranking method, the time to conduct the relevance ranking only increases linearly with the number of documents in the collection, which makes the proposed context-enhanced document ranking method computationally efficient and suitable for supporting IR in the TPER domain. Based on the TPER semantic model, 201 semantic concept indexes were developed to represent all the document context concepts, and cover all the important information-seeking tasks of the process (ICT 2014), such as making stakeholder involvement plan, developing reasonable alternatives, and preparing environmental documents. The efforts to develop all semantic concept indexes largely depends on the complexity of the domain and the number of the concepts in the semantic model (Simperl and Mochol 2006). The TPER process has moderate complexity – because it has been explicitly defined by numerous regulations and guidance documents – and the TPER semantic model has a medium number of concepts (201 concepts), which makes the development of the semantic context indexes relatively efficient.

5.3.6 Error Analysis

One main type of error was identified based on the retrieval results. Documents that have unbalanced semantic similarities to the semantic query (high semantic similarities to some concepts but low or zero semantic similarities to other concepts) were unfairly given high relevance scores. This is because each concept has the same contribution to the relevance score based on the assumption that a user gives equal importance to each concept in the query. For example, the query "environmental screening method for highway project in Florida" includes three concepts: "highway project", "environmental screen process", and "Florida". The document shown in Figure 5.4 describes the environmental screening tool for Colorado, and was mistakenly retrieved as one of the top 10 relevant documents because it has high semantic similarities to the concept "highway project" and "environmental screen process" but low semantic similarity to the concept "Florida". In future work, the semantic relevance ranking method could be improved by penalizing documents with unbalanced semantic similarities and/or by implementing different weights for each concept during relevance evaluation.



GEOGRAPHIC INFORMATION SYSTEM & ENVIRONMENTAL SCREENING TOOL
Program Description
The Colorado Department of Transportation (CDOT) has developed a new "Planning and Environmental Linkages (PEL) Decision Tool." The purpose of this tool is to identify red flags, including problematic historic properties, during transportation planning. This tool can be used by anyone within the DOT and by local planning organizations that are provided access to the PEL tool through CDOT. The tool consists of a series of information and decision-making screens, including a screen on "Archaeological/Historical/Paleontological" resources. The final output using this tool is an environmental planning report.
Program Elements
Historic Preservation in Transportation Planning, Historic Preservation in Early Project Development, Interagency Cooperation and Collaboration, Geographic Information Systems, Historic Property Database, Historic Property Screening Tool, Historic Property "Red Flags
Challenges Encountered
One initial challenge was funding. Guidance included Appendix A of 23 CFR 450 (the Appendix dealt with linking the transportation planning and National Environmental Policy Act processes) was used, in part, to justify the creation of the PEL program. Also, CDOT already was committed to similar "big picture" planning efforts, such as conservation banking. The PEL program was promoted as one of these planning efforts……

Figure 5.4 – A Document (Partial) Mistakenly Retrieved as one of the Top 10 Relevant Documents for a Testing Query

CHAPTER 6: STAKEHOLDER OPINION EXTRACTION FOR SUPPORTING ASPECT-LEVEL STAKEHOLDER OPINION MINING

6.1 State of the Art and Knowledge Gaps

The ML-based approach was used in this research because fine-grained and precise information is necessary for transportation project decision making. Some important research efforts, in the computer science domain, have been conducted to develop supervised ML-based methods for stakeholder opinion extraction. For example, Li et al. (2010) developed a new conditional random fields (CRF)-based framework to jointly extract positive opinions, negative opinions, and aspects from movie and product reviews, and proposed a skip-tree CRF algorithm to integrate the conjunction structure and the syntactic-tree structure. Shariaty and Moghaddam (2011) employed CRF to extract product aspects, corresponding opinions, and related usages from user reviews, and proposed techniques to solve the feature sparsity problem, conduct feature selection, and reduce the negative effect of excessive background words. Yang and Cardie (2012) proposed a semi-CRF-based approach to extract explicit and implicit opinion expressions from news articles, which takes the syntactic structure information into account during learning and inference, and identifies the opinion expressions at the segment level. Alghunaim et al. (2015) proposed the use of vector-based features computed by word-vector representations for extracting aspect terms from restaurant reviews, and applied the proposed features in two effective information extraction algorithms, CRF and SVM-HMM. Katiyar and Cardie (2016) investigated the use of deep bi-directional long short-term memory (LSTM) networks to jointly extract opinion entities and the relations that connect them, and improved the extraction performance through incorporating sentence-level and relation-level optimization.

In recent years, a number of research studies have been conducted on applying texting mining techniques in the construction domain. For example, Yu and Hsu (2012) proposed a content-based text mining technique to extract the textual content from a computer-aided design (CAD) document, and represented the textual content using a vector space model (VSM) to enable the automated and expedited retrieval of CAD documents based on similarity matching. Williams and Gong (2013) applied data mining classification algorithms to predict the level of cost overrun based on text descriptions of a project's characteristics and numerical data, such as the number of bidders and the low-bid price. Alsubaey et al. (2015) proposed a Naïve Bayes text mining approach to identify early warnings of project failures based on critical management documents such as minutes of meetings, and focused on identifying the warnings from project management aspects. To better understand the public interests in infrastructure projects, Nik-Bakht and El-Diraby (2015) utilized a K-means clustering algorithm to group the followers of the Toronto Light Rail Transit (LRT) mega project on Twitter based on semantic similarities among their user profiles, and analyzed the project-related tweets of each group using latent semantic indexing (LSI) to find the public interest topics.

Despite the importance of the aforementioned research efforts, there is no method that can effectively extract stakeholder opinions from stakeholder comments on large-scale transportation projects to support transportation decision making. On one hand, existing efforts that focused on the infrastructure domain cannot extract precise or fine-grained aspects. For example, although Nik-Bakht and El-Diraby's (2015) focused on analyzing tweets about infrastructure projects, they adopted a topic modeling-based method to identify public interest topics, which is not suitable for finding precise and fine-grained stakeholder concerns. On the other hand, existing research efforts on stakeholder opinion extraction mostly focused on product or service reviews, which are very different from stakeholder opinions on large-scale transportation projects in terms of the opinions and the concerns expressed, and the linguistic patterns displayed. Because text features in stakeholder comments can vary from one domain to another (e.g., product reviews versus transportation project reviews), it is very difficult for a single information extraction method to produce equally reliable performance results across different domains (Jakob and Gurevych 2010; Li et al. 2012; Chen et al. 2014). A transportation-project-domain-specific stakeholder information extraction method is, thus, needed for extracting stakeholder opinions from stakeholder comments on large-scale transportation projects.

In this regard, three main knowledge gaps have been identified. First, existing stakeholder opinion extraction methods cannot extract subject (opinion target), concern (aspect), and opinion expressions at the same time; they either focused on extracting either aspect or opinion expressions only [e.g., Yang and Cardie (2012) and Alghunaim et al. (2015)], or focused on extracting both aspect and opinion expressions but without extracting the opinion target expressions [e.g., Li et al. (2010) and Shariaty and Moghaddam (2011)]. Because stakeholder opinions on transportation projects have a finer level of opinion targets such as different design alternatives and route options, it is necessary to extract these opinion target expressions to better support transportation decision making. Second, the impact of dependency features and semantic features, especially domain-specific semantic features, on the performance of stakeholder opinion extraction has not been comprehensively evaluated. For example, Shariaty and Moghaddam (2011) only evaluated the performance of syntactic features, and Yang and Cardie (2012) did not evaluate the impact of dependency features. Third, there is lack of efforts to improve the recall of existing ML-based stakeholder opinion extraction methods. To ensure stakeholder concerns and support levels can be identified from the extracted opinion information in a complete and accurate manner, the opinion extraction method should achieve high performance in terms of both precision and recall. However, most of the existing ML-based stakeholder opinion extraction methods gave insufficient recall performance. For example, the best performance method proposed by Alghunaim et al. (2015) achieved 82.7% precision, but only 74.2% recall. It is thus important to develop strategies and methods that improve the recall of opinion extraction.

6.2 Proposed Stakeholder Opinion Extraction Method

To address the aforementioned knowledge gaps, this research proposes a stakeholder opinion extraction methodology, which extracts subject, concern, and opinion expressions from stakeholder comments on large-scale transportation projects. The stakeholder opinion extraction methodology is summarized in Figure 6.1. A stakeholder concern is an issue that is affected, positively or negatively, by the project, such as property value, farmland, fuel tax, population growth, and nearby environmental resources. A concern expression is a word or phrase that expresses a stakeholder concern. A subject expression is a word or phrase that refers to the target object of the comment, such as a project or an element of the project such as a design alternative or a route selection. An opinion expression is a word or phrase that expresses the opinions of a stakeholder. In developing the stakeholder opinion extraction methodology, the performances of five ML algorithms were evaluated: HMM, MEMM, CRF, SP, and SVM-HMM.

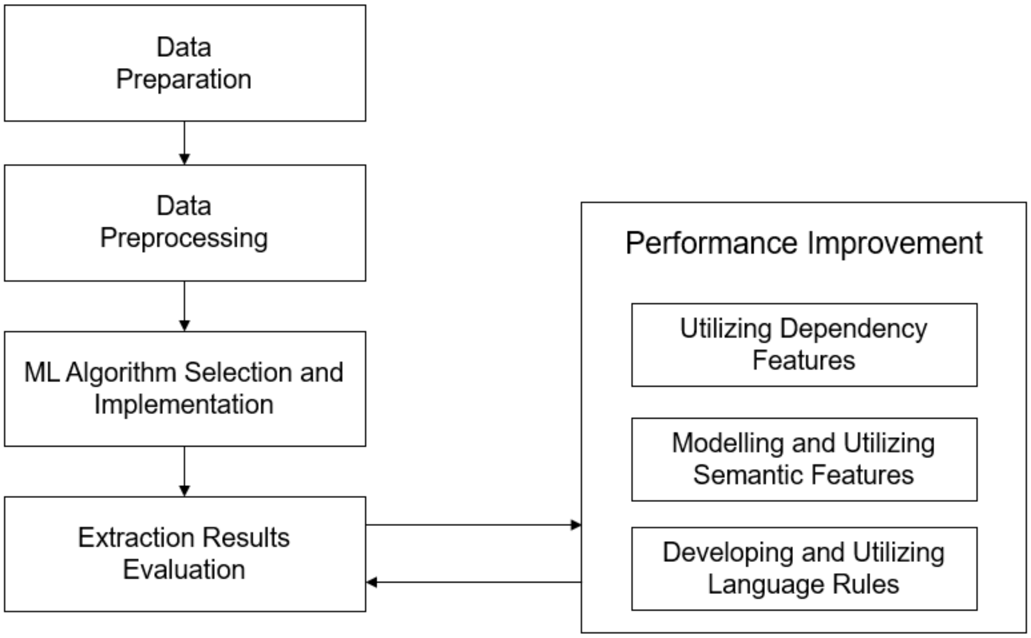

Data Preparation
Data Preprocessing
ML Algorithm Selection and Implementation
Extraction Results Evaluation
Performance Improvement
Utilizing Dependency Features
Modelling and Utilizing Semantic Features
Developing and Utilizing Language Rules

Figure 6.1 – The Proposed Stakeholder Opinion Extraction Methodology

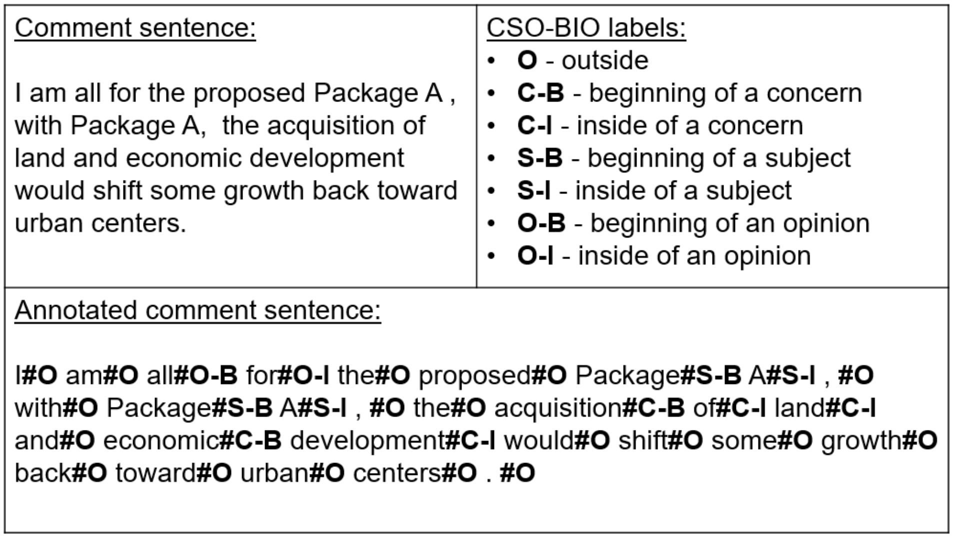
6.2.1 Data Preparation

To create a comment collection, nine large-scale transportation projects from nine states were identified (Table 6.1). The projects were selected from different geographic locations across the country, in order to have coverage of different project subjects, stakeholder concerns, and opinions in the collection. For these projects, the comments from all stakeholder groups (federal agencies, state agencies, local governments, public organizations, and interested individuals) that were received during the public comment period – including comments submitted through the project websites, emails, and public hearings – were gathered into a comment collection. As shown in Table 6.1, the comment collection contains 3,112 comments with a total number of 22,222 sentences.

Table 6.1 – Statistics of the Collected Comments

Project name	Project location	Number of comments	Number of sentences	Average sentences per comment
Cleveland Opportunity Corridor	Ohio	136	394	3
I-395 Transportation System	Maine	134	404	3
Illiana Corridor Tier 1	Illinois & Indiana	1,122	8,560	8
OR62 Corridor	Oregon	64	407	6
US281	Texas	641	5,725	9
Crosstown Parkway	Florida	35	333	10
Gulf Coast Parkway	Florida	42	345	8
I-5	California	339	2096	6
North I-25	Colorado	599	3958	7
Total	NA	3,112	22,222	7

A total of 500 comments were randomly selected from the comment collection – 400 for training and 100 for testing, which include 1,823 and 440 sentences, respectively. To create the gold standards, both the training and the testing datasets were annotated by three annotators (the author and two other researchers). The Begin-Inside-Outside (BIO) labeling schema was adopted to annotate each term of a comment sentence, while considering the type of expression (concern, subject, or opinion). The adapted schema was, thus, called the concern-subject-opinion (CSO)-BIO labeling schema]. The CSO-BIO labels "C-B", "S-B", and "O-B" indicate that the term is the beginning of a concern expression, subject expression, and opinion expression, respectively; "C-I", "S-I","O-I" indicate that the term is inside a concern expression, subject expression, and opinion expression, respectively; and "O" indicates the term is not a part of either a subject, concern, or opinion expression. For example, the expression "acquisition of land" is a concern expression about land use, which was annotated as "acquisition#C-B of#C-I land#C-I". Figure 6.2 provides an example of an annotated sentence from the comment collection.



Comment sentence: I am all for the proposed Package A , with Package A, the acquisition of land and economic development would shift some growth back toward urban centers.	CSO-BIO labels: • **O** - outside • **C-B** - beginning of a concern • **C-I** - inside of a concern • **S-B** - beginning of a subject • **S-I** - inside of a subject • **O-B** - beginning of an opinion • **O-I** - inside of an opinion

Annotated comment sentence:

I**#O** am**#O** all**#O-B** for**#O-I** the**#O** proposed**#O** Package**#S-B** A**#S-I** , **#O** with**#O** Package**#S-B** A**#S-I** , **#O** the**#O** acquisition**#C-B** of**#C-I** land**#C-I** and**#O** economic**#C-B** development**#C-I** would**#O** shift**#O** some**#O** growth**#O** back**#O** toward**#O** urban**#O** centers**#O** . **#O**

Figure 6.2 – An Example of an Annotated Sentence from the Comment Collection

6.2.2 Data Preprocessing

Data preprocessing is the process of preparing the comments in the training and testing datasets for further processing and machine learning. Each term in a comment sentence was transferred into a fixed-size feature vector. In the baseline case, only syntactic features were considered.

Syntactic features characterize the syntactic attributes of the terms. Four types of features were used in this research: tokens, part-of-speech (POS) tags, lemmas, and stopwords. Tokens are meaningful elements that form a sentence such as words, phrases, or symbols. In this research, a single word or a punctuation was regarded as a common token. Punctuations were not removed because they are natural boundaries of phrases and sentences, which can provide useful information to better identify the desired subject, concern, and opinion expressions. A POS tag defines the syntactic function of a word in a sentence such as noun, adjective, and verb. The Stanford POS Tagger (Manning et al. 2014) was used for POS tagging. A lemma is the dictionary form of a term, and is obtained through removing the inflectional ending of the term. For example, the words “opposes”, “opposed”, and “opposing” would have the same lemma “oppose”. The stopword feature defines whether the current term is a stropword or not. Stopwords are those words that have high frequency but low discriminating power, such as “to” “in”, “on”, and “the”. Although stopwords are often removed for common natural language processing tasks, they were retained in this research for two reasons. First, the subject, concern, and opinion expressions may contain stopwords. Second, stopwords can be good indicators of the desired expressions. For example, in the comment sentence “I am all for the toll road going in”, the opinion expression “all for” contains the stopwords “all” and “for”, and the subject expression “toll road” is following the stopword “the”.

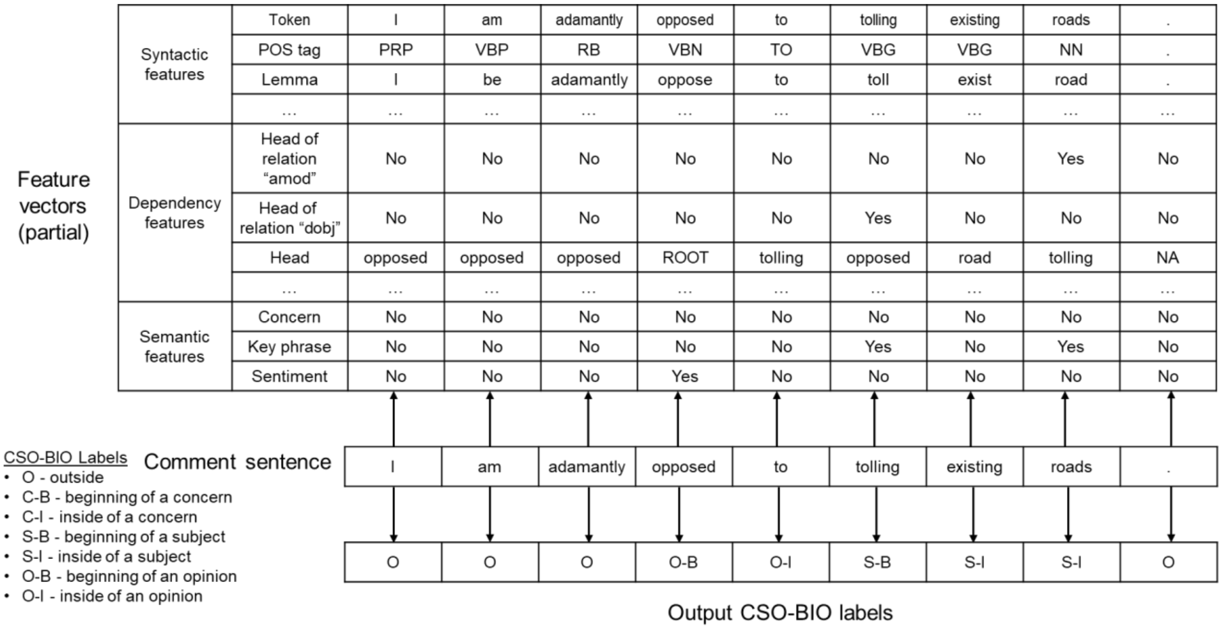
The syntactic information (tokens, POS tags, lemmas, and stopwords) of the surrounding four terms (including two terms occurring before the current term, and two terms occurring after the current term) were considered as part of the syntactic features, for two reasons. First, the syntactic information of the surrounding terms could affect the label of the current term, because the subject, concern, and opinion expressions can contain more than one term, in case these expressions are multi-term phrases. Second, the majority of the subject, concern, and opinion expressions in the training and testing data have less than five terms.

6.2.3 Machine Learning Algorithm Implementation and Testing

The subject, concern, and opinion extraction task was formulated as a sequence labeling task, which aims to assign a categorical label (i.e., an CSO-BIO label) to each member of the observation sequence (i.e., each term or punctuation in a comment sentence). In the context of this research, each comment sentence was preprocessed as a sequence of feature vectors, where the target output is the label of each term. Figure 6.3 shows an example of the partial feature vectors and the output CSO-BIO labels for a comment sentence.

A set of supervised machine learning algorithms that are commonly used for sequence labeling tasks were implemented and tested (using the same syntactic features): HMM, MEMM, linear-chain CRF, SP, and SVM-HMM (implemented in linear kernel). Although numerous studies have applied the above-mentioned algorithms for information extraction, no study has evaluated all the above algorithms on extracting stakeholder opinions on highway projects. Because text features in comments can vary from one domain to another, thus leading to variance in the information extraction performance, it is important to compare the performance of these algorithms when developing stakeholder opinion extraction method in the highway project domain. For HMM, a combination of different syntactic features of the words are identified as the observations, and the opinion expression labels as the underlying states. The transition probabilities from a previous label to a current label was estimated based on their occurrences in the training data, and unfairly favored the transition from label “O” to the same label “O” over other labels (“C-B”, “S-B”, or “O-B”). For the experiments conducted in this research, a linear-chain CRF algorithm and a linear kernel SVM-HMM algorithm were implemented.

Each algorithm has some important parameters that were tuned and optimized through trial and error based on the extraction performance. For example, parameter *C* in SVM-HMM controls the trade-offs between tolerance for mislabeling and the complexity of the model, which can significantly affect the performance of the algorithm when labeling unseen terms. To optimize the parameter *C* in SVM-HMM, first, an initial set of values (e.g., 0.001, 0.01, 0.1, 1, 10, and 100) were evaluated to identify the approximate magnitude of *C*. Then, a range of specific values (e.g., 1 to 10 at 0.1 interval) in that magnitude was evaluated to find the value of *C* that has the best performance. The HMM and SP algorithms were implemented using the Seqlearn sequence classification library for Python (Buitinck 2014); the MEMM algorithm was implemented using the sequence labelling toolkit Wapiti (Lavergne et al. 2010); the linear-chain CRF algorithm was implemented using the sklearn-crfsuite package (Korobov 2015), a python wrapper for the CRFsuite toolkit (Okazaki 2007); and the SVM-HMM was implemented using the sequence tagging with structural SVM package (Joachims 2008).



Feature vectors (partial)

	Token	I	am	adamantly	opposed	to	tolling	existing	roads	.
Syntactic features	POS tag	PRP	VBP	RB	VBN	TO	VBG	VBG	NN	.
	Lemma	I	be	adamantly	oppose	to	toll	exist	road	.
	…	…	…	…	…	…	…	…	…	…
Dependency features	Head of relation "amod"	No	No	No	No	No	No	No	Yes	No
	Head of relation "dobj"	No	No	No	No	No	Yes	No	No	No
	Head	opposed	opposed	opposed	ROOT	tolling	opposed	road	tolling	NA
	…	…	…	…	…	…	…	…	…	…
Semantic features	Concern	No	No	No	No	No	No	No	No	No
	Key phrase	No	No	No	No	No	Yes	No	Yes	No
	Sentiment	No	No	No	Yes	No	No	No	No	No

Comment sentence	I	am	adamantly	opposed	to	tolling	existing	roads	.
Output CSO-BIO labels	O	O	O	O-B	O-I	S-B	S-I	S-I	O

CSO-BIO Labels
- O - outside
- C-B - beginning of a concern
- C-I - inside of a concern
- S-B - beginning of a subject
- S-I - inside of a subject
- O-B - beginning of an opinion
- O-I - inside of an opinion

Figure 6.3 – An Example of the Partial Feature Vectors and the Output CSO-BIO Labels for a Comment Sentence

6.2.4 Evaluation

The performance of the developed algorithms was evaluated using precision, recall, and F1 measure, as per Eqs. (6.1), (6.2), and (6.3), where true positive (*TP*) refers to the number of opinion expressions extracted correctly, false positive (*FP*) refers to the number of opinion expressions extracted incorrectly, and false negative (*FN*) refers to the number of opinion expressions incorrectly extracted as negative. Precision, here, is defined as the ratio of the number of correctly extracted expressions over the total number of extracted expressions. Recall, here, is defined as the ratio of the number of correctly extracted expressions over the total number of expressions that should be extracted. F1 is the harmonic mean of precision and recall. These measures were calculated based on a comparison of the extraction results with the gold standard annotations.

$$Precision = \frac{TP}{TP+FP} \tag{6.1}$$

$$Recall = \frac{TP}{TP+FN} \tag{6.2}$$

$$F1\ measure = \frac{2*Precision*Recall}{Precision+Recall} \tag{6.3}$$

6.2.5 Proposed Methods for Improving the Performance of Opinion Extraction

Three methods were proposed to improve the performance of opinion extraction: (1) utilizing dependency features, (2) modeling and utilizing semantic features, including two domain-specific semantic features; and (3) developing and utilizing a set of language rules, based on linguistic patterns.

6.2.5.1 Utilizing Dependency Features

Dependency features were utilized to capture the syntactic relations between the terms. Dependency features use the information about the dependency relations in a comment sentence. In a sentence, linguistic units, such as words and phrases, are connected to each other by dependency relations, which are grammatical relations between a head and a dependent. This feature group includes four different types of features: relation head, relation dependent, head, and POS of the head. The relation head feature represents whether the current term is the head of the selected dependency relations. The relation dependent feature represents whether the current term is the dependent of the selected dependency relations. The head feature represents the head of the current term in the dependency tree. The POS of the head feature represents the POS tag of the head term. Each comment sentence in the training and testing dataset was parsed by the Stanford dependency parser (Manning et al. 2014) to capture the dependency features of each term in the sentence.

Four different dependency relations were considered: the adjectival modifier relation “amod”, the determiner relation “det”, the direct object relation “dobj”, and the nominal subject relation “nsubj”. For example, for the comment sentence “the preferred alternative would likely impact wetlands within and connected to Midewin”, the four term pairs that have the aforementioned dependency relations are as follows: “alternative” and “preferred” have the adjectival modifier relation “amod”, where “alternative” is the head term and “preferred” is the dependent term; “alternative” and “the” have the determiner relation “det”, where “alternative” is the head term and “the” is the dependent term; “impact” and “wetland” have the direct object relation “dobj”, where “impact” is the head term and “wetland” is the dependent term; and “impact” and “alternative” have the nominal subject relation “nsubj”, where “impact” is the head “term” and “alternative” is the dependent term.

During the training process of the linear-chain CRF, all the input features were transformed into binary features, resulting in a large increase to the total number of features. The Elastic Net (L1 + L2) regularization (Zou and Hastie 2005) was then used to prevent over-fitting and conduct implicit feature selection, since the L1 regularization removes the non-effective features by assigning their parameters to zero (Ng 2004). For example, the total number of syntactic and dependency features (after transformation) was 67,578, but after training with the Elastic Net regularization the number of effective features was reduced to 13,462.

6.2.5.2 Modeling and Utilizing Semantic Features

Three models were utilized to capture the semantic features of the text, in order to further improve the performance of the information extraction algorithm: a stakeholder concern hierarchy, a key phrase list, and a sentiment lexicon. Accordingly, three types of semantic features were defined and used: concern features, key phrase features, and sentiment features.

The concern feature is a domain-specific feature that represents whether a term belongs to a concept in the stakeholder concern hierarchy. As mentioned in Chapter 4, the stakeholder concern hierarchy – which is part of the TPER Epistemology – was developed based on a literature review on transportation decision making and stakeholder involvement processes, and interactions with transportation practitioners during one-to-one interviews. The most abstract concept in the hierarchy is the “stakeholder concern”, which includes five main subconcepts: “environmental concern”, “transportation concern”, “socio-economic concern”, “cultural concern”, and “management concern”. These subconcepts were further decomposed, forming a hierarchy, at a total of 123 concepts. All 123 concepts were included in a stakeholder concern name list. The lemma of each term in a comment sentence was compared with the concepts in the stakeholder concern list to find out whether the term belongs to a concept in the stakeholder concern hierarchy. A partial view of the stakeholder concern hierarchy is shown in Figure 6.4.

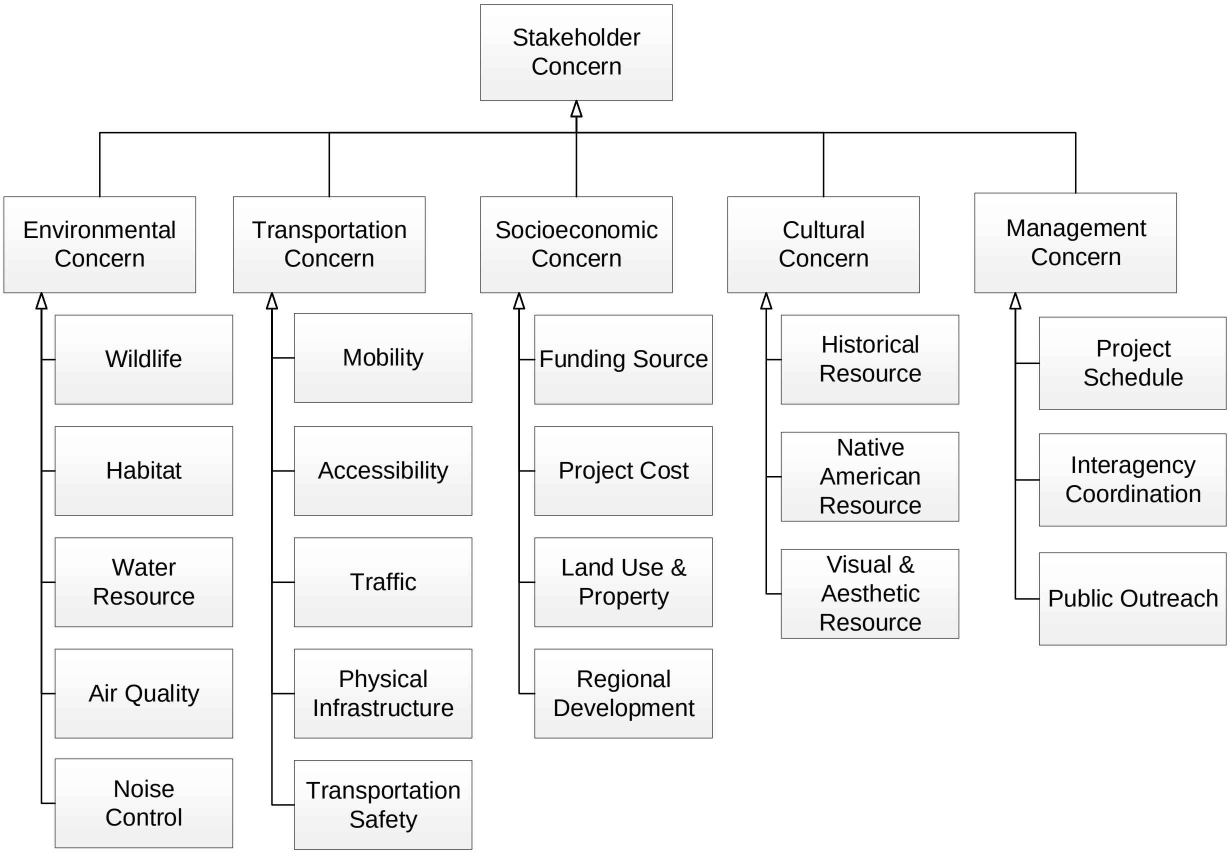


Figure 6.4 – A Partial View of the Stakeholder Concern Hierarchy

The key phrase feature is a domain-specific feature that represents whether a term is part of a key phrase in the key phrase list. Because concern expressions and subject expressions are mostly noun phrases that are frequently mentioned in the stakeholders' comments, a key phrase list including all noun phrases that have three or more appearances in the comment collection (excluding training and testing data) was developed. Each comment sentence was first parsed by the Stanford NLP parser (Manning et al. 2014) to obtain all the appearing noun phrases, and the frequency of each extracted phrase in the whole comment collection was then determined. All the extracted phrases that appeared less than three times in the comment collection were deleted, and the remaining phrases were included in the key phrase list. The lemma of each term in a comment sentence was compared with the key phrases in the key phrase list to find out whether the term is part of a key phrase.

The sentiment feature is a general feature that represents whether a term is a positive word or a negative word in the sentiment lexicon. The sentiment lexicon by Hu and Liu (2004) was utilized, which includes 2,006 positive opinion words and 4,780 negative opinion words. The lemma of each term in a comment sentence was compared with the terms in the sentiment lexicon to find out whether the term is a positive word or a negative word.

6.2.5.3 Developing and Utilizing a Set of Language Rules

To further improve the recall of the information extraction algorithms, a set of language rules were developed based on the analysis of the linguistic patterns displayed in the stakeholder comment collection.

- Rule R1: If the nominal subject of a verb or verb phrase is a subject expression, then the direct object of the verb or verb phrase is a concern expression, and vice versa. For example, in the comment "the A3S2 working alignment would impact approximately 10.3 acres of forested land", the nominal subject of the verb "impact" is "the A3S2 working alignment", which is a subject expression, then the direct object "10.3 acres of forested land" would be labeled as a concern expression.
- Rule R2: If a noun or verb phrase is in conjunction with another concern expression, then the noun or verb phrase is also a concern expression. For example, in the comment "a train will be better for the air quality and for seniors, physically handicapped, and others who cannot drive on I-25", the concern expression "air quality" is in conjunction with three

other noun phrases: “seniors”, “physically handicapped”, and “others who cannot drive on I-25”. These three noun phrases would be labeled as concern expressions.

- Rule R3: If the nominal subject of a copular verb (such as be, are, appear) is a subject expression, then the adjective or the noun compliment that follows the copular verb is an opinion expression, and vice versa. For example, in the comment “but the southern route would be my first choice”, the nominal subject of the copular verb “be” is “the southern route”, which is a subject expression, then the noun compliment “my first choice” would be labeled as an opinion expression.
- Rule R4: If the direct object of a verb or verb phrase that follows a personal pronoun (such as I and we) is a subject expression, then the verb or verb phrase is an opinion expression, and vice versa. For example, in the comment “ I am strongly opposed the widening of the I-5”, the verb phrase “strongly opposed” follows a personal pronoun “I”, and has the direct object “widening of I-5”, which is a subject expression, then the verb phrase “strongly opposed” would be labeled as an opinion expression.

The language rules and the selected machine learning algorithm were combined as follows. First, the machine learning algorithm was used to extract the initial subject, concern, and opinion expressions. Then, the language rules were applied to each comment sentence, iteratively, until there were no new subject, concern, or opinion expressions extracted.

6.3 Experimental Results and Analysis

A number of experiments were conducted to (1) evaluate and select the supervised machine learning algorithm that yields the best performance for stakeholder opinion extraction; and (2) evaluate and demonstrate the effects of dependency features, semantic features, and the language rules on the performance of the extraction. The final combination of the methods that were selected for all steps forms the proposed information extraction method.

6.3.1 Performance of Different Machine Learning Algorithms

The performance of the five algorithms are summarized in Table 6.2. As shown in Table 6.2, the performances of concern and subject extraction are much better than the performance of opinion extraction for every machine learning algorithm implemented. While the concern and subject expressions are mostly noun phrases, opinion expressions can take a number of different forms, such as noun phrase (“first choice” in the comment “but the southern route would be my first choice”), verb phrase (“opposed to” in the comment “I’m vehemently opposed to this road going in”), adjective phrase (“impractical, cumbersome, and most of all ill-advised” in the comment “tolls are impractical, cumbersome, and most of all ill-advised for solving costs and traffic”), or prepositional phrase (“in support of” in the comment “I am in support of the DEIS package A”). Compared with concern and subject expressions, opinion expressions are thus much more difficult to capture.

For the subject extraction, the linear-chain CRF algorithm achieved the best precision of 92%, the second-best recall of 79%, and the best F1 measure of 85%, among the five implemented algorithms. For the concern extraction, the linear-chain CRF algorithm achieved the second-best precision of 92%, the best recall of 80%, and the best F1 measure of 85%. For opinion expression extraction, the linear-chain CRF achieved the best precision of 76%, the second-best recall of 63%, and the best F1 measure of 69%. Overall, the linear-chain CRF achieved the second-best precision 89%, the second-best recall of 76%, and the best F1 measure of 82%. Based on the F1 measure performance on all the three extraction tasks, the linear-chain CRF algorithm was selected to further implement and test the performance of improvement strategies.

Table 6.2 – Information Extraction Performance of the Five Machine Leaning (ML) Algorithms

ML Algorithm	Subject extraction			Concern extraction			Opinion extraction			Overall		
	P*	R*	F1	P*	R*	F1	P*	R*	F1	P*	R*	F1
HMM	76%	63%	69%	74%	64%	69%	62%	44%	52%	73%	60%	66%
MEMM	78%	66%	71%	80%	67%	73%	64%	50%	56%	76%	63%	69%
Linear-chain CRF	92%	79%	85%	92%	80%	85%	76%	63%	69%	89%	76%	82%
SP	84%	83%	83%	82%	80%	81%	63%	64%	64%	79%	78%	78%
SVM-HMM	92%	72%	81%	93%	67%	78%	68%	28%	39%	90%	61%	73%

* P=precision; R=recall

6.3.2 Effect of Utilizing Dependency Features

As shown in Table 6.3, after adding the dependency features, for subject expression extraction, the precision was improved from 92% to 95%, the recall was improved from 79% to 80%, and the F1 measure was improved from 85% to 87%. For concern expression extraction, the precision was improved from 92% to 94%, the recall was improved from 80% to 81%, and the F1 measure was improved from 85% to 87%. For opinion expression extraction, the precision was improved from 76% to 77%, the recall was unchanged, and the F1 measure was improved from 69% to 70%. Because most of the dependency features generated using the four dependency relations (“amod”, “det”, “dobj”, and “nsubj”) are good indicators of noun phrases, opinion extraction did not receive an equal improvement in performance (1% increase in F1 measure compared with 2% increase for both subject and concern extraction) due to its more complicated nature. Overall, the precision was improved from 89% to 92%, the recall was improved from 76% to 77%, and the F1 measure was improved from 82% to 84%.

Table 6.3 – Information Extraction Performance Improvement Using Dependency Features

Features	Subject extraction			Concern extraction			Opinion extraction			Overall		
	P*	R*	F1	P*	R*	F1	P*	R*	F1	P*	R*	F1
Syntactic features	92%	79%	85%	92%	80%	85%	76%	63%	69%	89%	76%	82%
Syntactic features + dependency features	95%	80%	87%	94%	81%	87%	77%	63%	70%	92%	77%	84%

* P=precision; R=recall

6.3.3 Effect of Utilizing the Proposed Semantic Features

After adding the three types of semantic features, the linear-chain CRF algorithm was able to identify the subject, concern, and opinion expressions that were otherwise not extracted. For example, in the comment “it’s simply ridiculous that we don't have an extensive, reliable commuter rail”, the opinion expression “ridiculous” was identified due to the use of the sentiment lexicon.

As shown in Table 6.4, after adding the semantic features (including concern feature, key phrase feature, and sentiment feature), for subject expression extraction, the precision was improved from 92% to 94%, the recall was improved from 79% to 81%, and the F1 measure was improved from 85% to 87%. For concern expression extraction, the precision was improved from 92% to 95%, the recall was improved from 80% to 81%, and the F1 measure was improved from 85% to 87%. For opinion expression extraction, the precision was improved from 76% to 88%, the recall was improved from 63% to 74%, and the F1 measure was improved from 69% to 80%. Compared with subject and concern extraction, opinion extraction showed the greatest performance improvement (11% increase in F1 measure) after adding the semantic features. This is largely due to the fact that the sentiment lexicon contains many opinion terms that are hard to identify using only syntactic and dependency features, such as opinion terms that have no appearance in the training data, and verbs or nouns that express positive or negative sentiments.

Using both the dependency and semantic features with the original syntactic features, for subject expression extraction, the precision was improved from 92% to 97%, the recall was improved from 79% to 81%, and the F1 measure was improved from 85% to 89%. For concern expression extraction, the precision was improved from 92% to 96%, the recall was improved from 80% to 81%, and the F1 measure was improved from 85% to 88%. For opinion expression extraction, the precision was improved from 76% to 92%, the recall was improved from 63% to 74%, and the F1 measure was improved from 69% to 82%. Overall, the linear-chain CRF algorithm with syntactic, dependency, and semantic features achieved 95% precision, 80% recall, and 87% F1 measure.

Because F1 measure integrates both precision and recall, their values were analyzed to evaluate whether the performance improvement after using both the dependency and semantic features is statistically significant. The Wilcoxon signed-rank test was used to examine whether the improvement in F1 measure is significant across the 10-fold cross validation results on the training data. The Wilcoxon signed-rank test is a nonparametric test for comparing the differences between two-paired samples (Rey and Neuhäuser 2011). The result of the Wilcoxon signed-rank test was interpreted according to the probability value (p-value). The p-value is 0.027, which is less than the 0.05 significance level. This indicates that there is a significant improvement in F1 measure when using both dependency and semantic features.

Table 6.4 – Information Extraction Performance Improvement Using Semantic Features

Features	Subject extraction			Concern extraction			Opinion extraction			Overall		
	P*	R*	F1	P*	R*	F1	P*	R*	F1	P*	R*	F1
Syntactic features	92%	79%	85%	92%	80%	85%	76%	63%	69%	89%	76%	82%
Syntactic features + semantic features	94%	81%	87%	95%	81%	87%	88%	74%	80%	93%	80%	86%
Syntactic features + dependency features + semantic features	97%	81%	89%	96%	81%	88%	92%	74%	82%	95%	80%	87%

* P=precision; R=recall

6.3.4 Performance of the Proposed Language Rules

Despite achieving 95% precision, the linear-chain CRF algorithm with syntactic, dependency, and semantic features achieved only 80% recall. To improve the recall of stakeholder opinion extraction, the set of developed language rules were combined with the linear-chain CRF algorithm and the feature combination. Using the language rules can greatly improve the recall by identifying the subject, concern, and opinion expressions that are not recognized by the linear-chain CRF. For example, for the comment "the 281 corridor needs more capacity", the linear-chain CRF only identified the concern expression “more capacity”, and applying the language rule R1 would extract the subject expression “281 corridor”. However, because the language rules may not work in every scenario, they could create errors that decrease the precision of information extraction. For example, in the comment “a national cemetery should not be impacted by highway safety concerns”, “highway safety concerns” was first extracted as a concern expression by the linear-chain CRF, but applying language rule R1 resulted in mistakenly labeling “a national cemetery” as a subject expression.

The performance results of the linear-chain CRF, alone and with language rules, are shown in Table 6.5. As shown in Table 6.5, after using language rules on the CRF-trained results, for subject expression extraction, although the precision dropped from 97% to 94%, the recall was improved from 81% to 88%, and the F1 measure was improved from 89% to 91%. For concern expression extraction, despite a 2% drop in precision (96% to 94%), the recall was improved from 81% to 89%, and the F1 measure was improved from 88% to 92%. For opinion expression extraction, although the precision dropped from 92% to 88%, the recall was improved from 74% to 88%, and the F1 measure was improved from 82% to 88%. Overall, using the CRF plus the language rules achieved 93% precision, 89% recall, and 91% F1 measure. To evaluate if the higher performance is statistically significant, the Wilcoxon signed-rank test was used to examine the improvements in F1 measure for all 10-fold cross validation results on the training data. The p-value for the F1 measures is 0.0044, which indicates that applying the language rules would significantly improve the performance of the opinion extraction.

As such, based on the experimental results and analysis, the proposed information extraction method is using linear chain CRF to extract the initial subject, concern, and opinion expressions based on the syntactic, dependency, and semantic features; and then iteratively applying the proposed language rules to improve the extraction performance.

Table 6.5 – Information Extraction Performance Improvement Using Language Rules

Algorithm	Subject extraction			Concern extraction			Opinion extraction			Overall		
	P*	R*	F1	P*	R*	F1	P*	R*	F1	P*	R*	F1
Linear-chain CRFs	97%	81%	89%	96%	81%	88%	92%	74%	82%	95%	80%	87%
Linear-chain CRFs + language rules	94%	89%	91%	94%	89%	92%	88%	88%	88%	93%	89%	91%

* P=precision; R=recall

6.3.5 Error Analysis

Three main types of errors were identified based on the testing results. First, irrelevant expressions were extracted as concern/subject/opinion expressions. For example, in the comment “a national cemetery should not be impacted by highway safety concerns”, “a national cemetery” was mistakenly extracted as a subject expression because of the language rule R1. To address this type of error, some strategies could be considered and tested in future work. For example, more sophisticated language rules could be developed to avoid similar errors. Second, relevant expressions were extracted with wrong labels (e.g., a concern expression was extracted as a subject expression, or vice versa). For example, in the comment “in general, we support bridging of wetlands rather than the placement of fill”, the phrase “bridging of wetlands” is supposed to be extracted as a subject expression, but was extracted as a concern expression. This is because the word “wetlands” appears most frequently as part of a concern expression in the training data. In future work, syntactic and semantic features could be incorporated to help recognize the right type of expression in cases where the same word appears in different types of expressions. Third, uncommon concern/subject/opinion expressions were not extracted. For example, in the comment “why should thousands of acres of farmland be impacted for a white elephant of an airport? “, the proposed method failed to extract “white elephant” as an opinion expression because the phrase did not appear in the training data nor in the sentiment lexicon. In future work, more comments could be collected and labeled as training data, and the stakeholder concern hierarchy and sentiment lexicon could be extended to cover uncommon concern/subject/opinion expressions.

CHAPTER 7: STAKEHOLDER OPINION CLASSIFICATION FOR SUPPORTING ASPECT-LEVEL STAKEHOLDER OPINION MINING

7.1 State of the Art and Knowledge Gaps

Opinion classification problems have long been studied in the field of opinion mining, with two main approaches that have been proposed in recent years: supervised approach and unsupervised approach. The supervised approach treats the opinion classification as a TC problem, and utilizes machine learning algorithms to classify stakeholder opinions through learning from labeled training data. For example, Gamallo and Garcia (2013) proposed a strategy based on a naïve Bayes (NB)-based classifier to classify tweets into two polarity categories: positive and negative. In addition to unigram features, they also incorporated n-gram phrases, built a polarity lexicon, and considered negative words that can shift the polarity of specific terms. Fang and Zhan (2015) compared the performance of three algorithms [NB, SVM, and random forest (RF)] on polarity categorization of online product reviews at both sentence and document level. They also proposed a negation phrase identification algorithm to incorporate negation phrases into semantic score computation, and used the semantic score as an important feature for classification.

Because sufficient labeled training data can be hard to obtain, many research studies have focused on using unsupervised approaches, which utilize topic models or lexicons to classify stakeholder opinions. For example, Lin and He (2009) developed a joint sentiment/topic model based on latent Dirichlet allocation (LDA) that can directly generate the probability distribution of a sentiment label given a document. Taboada et al. (2011) developed a semantic orientation calculator (SO-CAL) to assign a positive or negative label to a piece of text. The SO-CAL automatically extracts sentiment-bearing words to calculate semantic orientation, and incorporates valence shifters such as intensifiers and negation. Fernández-Gavilanes et al. (2016) proposed an unsupervised method that aimed to improve the sentiment classification performance through capturing syntactic and dependency information using natural language processing (NLP) techniques. They adapted the PolarityRank algorithm to create a contextualized sentiment lexicon from a set of positive and negative seed words, and propagated the term-level sentiments based on the syntactic structure of a sentence to predict the sentiment of the sentence.

Although a number of opinion classification studies have been conducted, there have been no research efforts for conducting aspect-level opinion classification in the infrastructure domain. Outside of the infrastructure domain, three primary knowledge gaps have been identified. First, most of the existing efforts rely on supervised machine learning algorithms, which require learning from a large amount of labeled training data to classify stakeholder opinions. On the other hand, existing efforts that took an unsupervised approach utilized algorithms such as topic-modeling, which are not suitable for classifying stakeholder opinions into precise and fine-grained concern categories to support highway decision making. Second, the majority of existing efforts focus only on sentiment classification, which is commonly solved as a binary or multiclass classification problem with a limited number of classes to assign (negative/positive or negative/neutral/positive), while concern classification is a multilabel classification problem with a greater number of classes and granularity levels. Third, there is a lack of comparison between unsupervised and supervised machine learning-based opinion classification approaches in terms of classification performance. For example, Fernández-Gavilanes et al. (2016) only compared their proposed unsupervised sentiment analysis method with other unsupervised methods, and Poria et al. (2016) only compared their proposed method with one supervised method on aspect extraction, but not on aspect-level sentiment classification.

7.2 Proposed Unsupervised Machine Learning-Based Opinion Classification Method

To address the aforementioned knowledge gaps, this research proposes an unsupervised machine learning-based aspect-level stakeholder opinion classification method, which can automatically create labeled training data through iteratively generating opinion tuple clusters based on keywords for each classification category. A supervised classifier then learns from the automatically-created training data to classify opinion tuples – extracted from stakeholder comments – into one or more concern categories and one sentiment category. The proposed method includes four primary elements (as per Figure 7.1): keyword identification, opinion characterization using semantic vectors, opinion tuple clustering, and opinion classification.

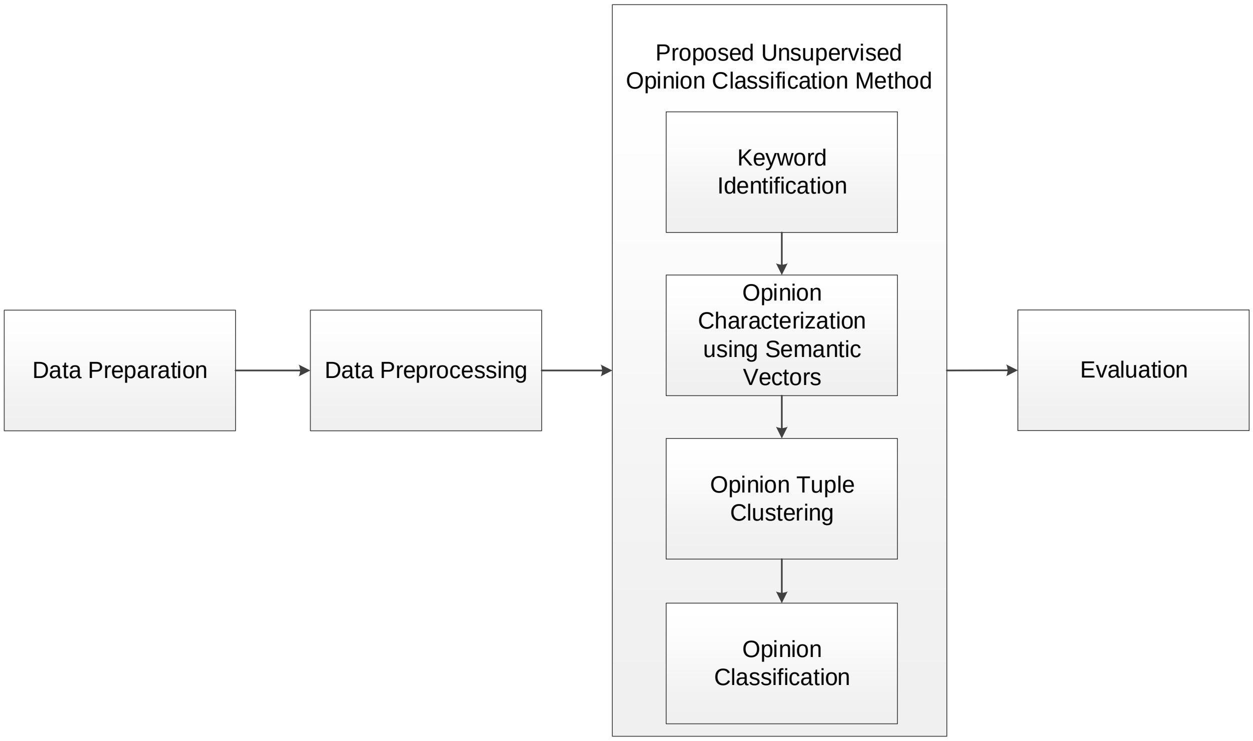


Figure 7.1 – Stakeholder Opinion Classification Methodology

7.2.1 Keyword Identification

A set of keywords are used to represent each category. For the concern categories, the keywords can be defined based on existing domain knowledge in the form of keyword lists, thesauri, taxonomies, ontologies, etc. For the sentiment categories, a sentiment lexicon – a domain-specific or general one – can be used. For example, for this research, a domain-specific keyword list was used for the concern categories. The keyword list was empirically developed based on the stakeholder concern hierarchy. An initial list of keywords was defined based on the terms in the names of the concern concepts and subconcepts and the synonyms of these terms. The final keywords were selected empirically. Figure 7.2 shows the stakeholder concern hierarchy, which includes the 14 categories that were used for classification. As mentioned in Chapter 4, the hierarchy was developed based on a literature review on transportation decision making and stakeholder involvement processes, and interactions with transportation practitioners during one-to-one interviews. Table 7.1 shows the final list of keywords, including a total of 31 keywords.

For the sentiment categories, which include three categories – supportive, unsupportive, and neutral, a set of keywords from a sentiment lexicon were used. The sentiment lexicon by Liu and Hu (2004), which includes 2,006 positive and 4,780 negative opinion words, was utilized. The positive and negative opinion words in the sentiment lexicon were used as keywords for the supportive and unsupportive categories, respectively.

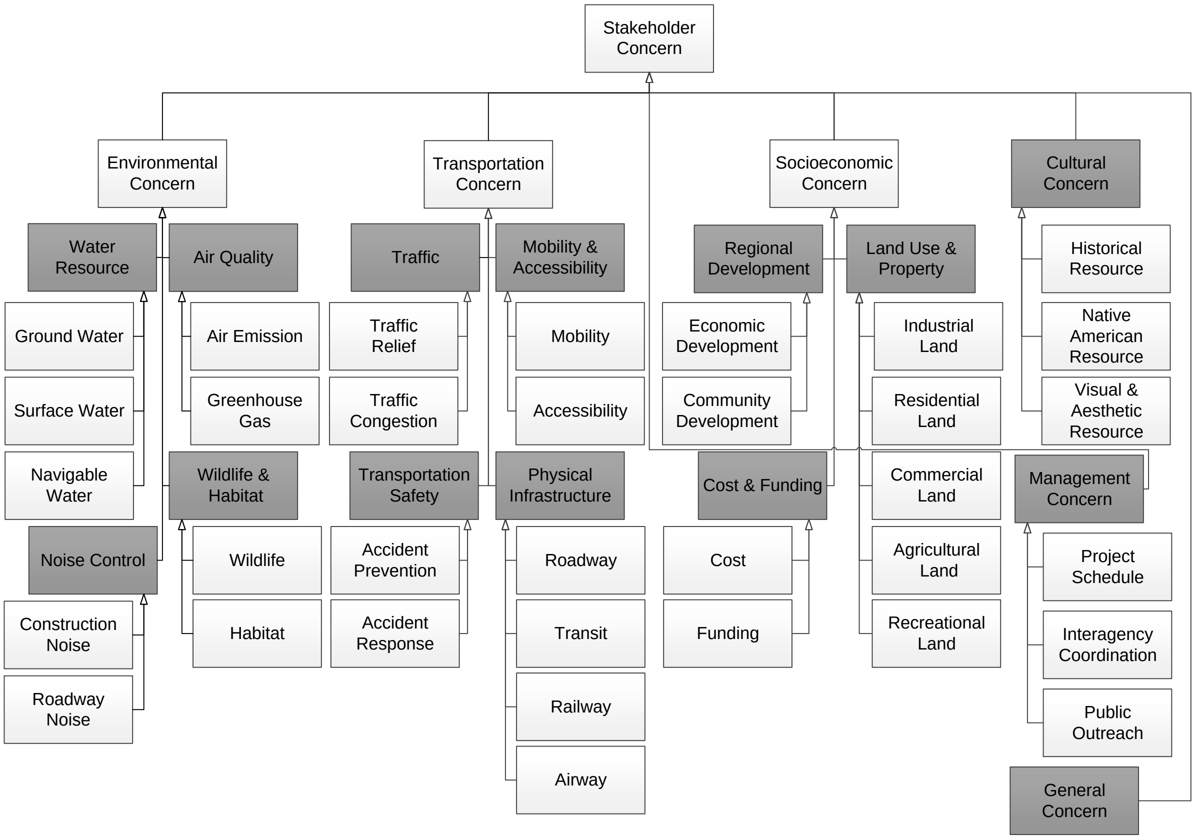


Figure 7.2 – A Partial View of the Stakeholder Concern Hierarchy

Table 7.1 – Sample Concern Category Keywords

Concern category	Keywords
Air quality	Air quality, air emission
Water resource	Water, wetland
Wildlife and habitat	Wildlife, habitat
Noise control	Noise, sound
Traffic	Traffic, congestion
Mobility and accessibility	Mobility, access
Physical infrastructure	Rail, bridge, overpass
Transportation safety	Accident, safety, safe
Cost and funding	Cost, fund
Land use and property	Land, property, home
Regional development	Economic, community, population
Cultural concern	Culture, aesthetic, historical
Management/administrative concern	Government, coordination

7.2.2 Opinion Characterization using Semantic Vectors

In order to develop opinion-tuple clusters for each classification category, opinion semantic vectors are proposed to capture the semantic similarities between opinion tuples. An opinion semantic vector is a real-valued vector of features that characterize the meaning of an opinion tuple, and is the weighted aggregation of the word semantic vectors of its words. A word semantic vector represents the contexts in which the word appears in a corpus of text.

In this research, four different types of opinion semantic vectors were developed and tested – using two different word-embedding models (Skip-gram and GloVe models) and two different corpuses (Wikipedia and highway stakeholder comment collection). The Ski-gram and GloVe models were selected because they are the state-of-the-art word-embedding models. They were tested because each offers different advantages to different tasks, and thus performs differently in different applications. For example, Levy et al. (2015) showed that the Skip-gram model outperformed the GloVe model in word similarity estimation, but Pennington et al. (2014) indicated that the GloVe model performed significantly better in word analogy assessment. Therefore, it is important to evaluate the performance of these two models when used to create opinion semantic vectors for opinion tuple clustering. Two different corpuses were tested to evaluate the impact of using opinion semantic vectors learned from a domain-specific corpus, here the highway stakeholder comment collection, on the classification performance.

The skip-gram and GloVe models were tested in generating the word embedding for each term in an opinion tuple. The skip-gram model (Mikolov et al. 2013) is a prediction-based model [a model that builds the word semantic vector through predicting the current word given the context words or vice versa (Baroni et al. 2014)]. It is a shallow, two-layer neural network that takes a word and its neighboring words within a context window as input, and predicts the probability for each word to actually appear in the context window. The model generates word embeddings from the weight matrix of the hidden layer. The GloVe model (Pennington et al. 2014) is a count-based model [a model that builds the word semantic vector based on word-cooccurrence statistics (Baroni et al.

2014)]. It is an unsupervised learning algorithm for obtaining vector representations for words, which performs training on aggregated global word-word cooccurrence statistics from a corpus, and learns word embeddings such that their dot product equals the logarithm of the words' cooccurrence probability. For both models, five was selected as the size of the context window, and 300 was the dimension of the vector.

The proposed opinion tuple O_i is represented in Eq. (7.1), where t_j is a term in the opinion tuple and *m* is the total number of terms in the opinion tuple.

$$O_i = \{\cup_{j=1}^{m} t_j\} \tag{7.1}$$

The opinion semantic vector V_i for the opinion tuple O_i is expressed in Eq. (7.2), where e_j is the word semantic vector for term t_j, w_j is the weight of term t_j, and *m* is the total number of terms in the opinion tuple O_i.

$$V_i = \{\cup_{j=1}^{m} w_j \, e_j\} \tag{7.2}$$

The weight w_j is proposed to accommodate terms with different discriminating powers and is defined in Eq. (7.3), where $TF(t_j)$ is the frequency of term t_j in the expression and $IEF(t_j)$ is the inverse expression frequency of term t_j.

$$w_j = TF(t_j) * IEF(t_j) \tag{7.3}$$

The inverse expression frequency of term t_j $IEF(t_j)$ is defined in Eq. (7.4), where N is the total number of expressions and $EF(t_j)$ is the number of expressions that contain the term t_j.

$$IEF(t_j) = \log(\frac{N}{EF(t_j)}) \tag{7.4}$$

7.2.3 Opinion Tuple Clustering

To create opinion tuple clusters for each classification category, the k-means clustering algorithm (Aggarwal and Reddy 2013) was adapted to incorporate the semantic similarities between opinion tuples and the characteristics of both concern and sentiment classification. Figures 7.3 and 7.4 provide examples of how opinion tuple clusters are generated for concern and sentiment classification. The adapted k-means clustering algorithm includes two steps: initialization step and iteration step.

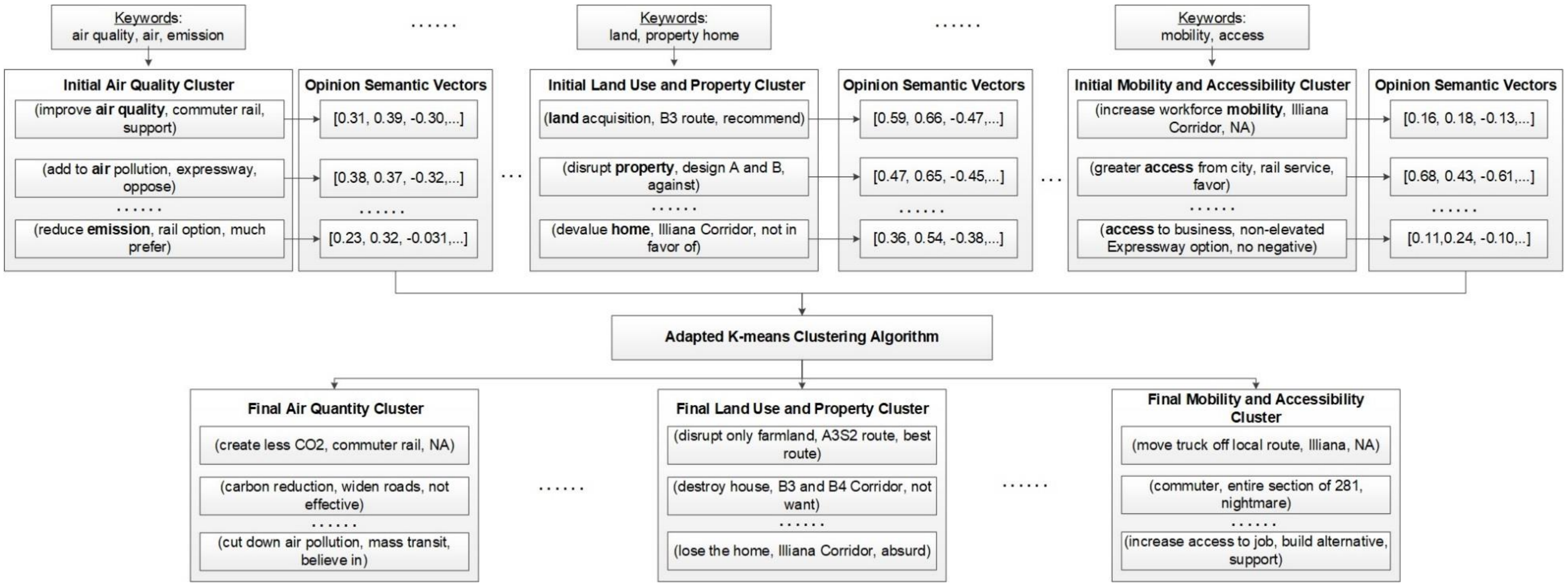

Keywords:
air quality, air, emission
......
Keywords:
land, property home
......
Keywords:
mobility, access
Initial Air Quality Cluster
(improve air quality, commuter rail, support)
(add to air pollution, expressway, oppose)
......
(reduce emission, rail option, much prefer)
Opinion Semantic Vectors
[0.31, 0.39, -0.30,...]
[0.38, 0.37, -0.32,...]
......
[0.23, 0.32, -0.031,...]
...
Initial Land Use and Property Cluster
(land acquisition, B3 route, recommend)
(disrupt property, design A and B, against)
......
(devalue home, Illiana Corridor, not in favor of)
Opinion Semantic Vectors
[0.59, 0.66, -0.47,...]
[0.47, 0.65, -0.45,...]
......
[0.36, 0.54, -0.38,...]
...
Initial Mobility and Accessibility Cluster
(increase workforce mobility, Illiana Corridor, NA)
(greater access from city, rail service, favor)
......
(access to business, non-elevated Expressway option, no negative)
Opinion Semantic Vectors
[0.16, 0.18, -0.13,...]
[0.68, 0.43, -0.61,...]
......
[0.11,0.24, -0.10,..]
Adapted K-means Clustering Algorithm
Final Air Quantity Cluster
(create less CO2, commuter rail, NA)
(carbon reduction, widen roads, not effective)
......
(cut down air pollution, mass transit, believe in)
......
Final Land Use and Property Cluster
(disrupt only farmland, A3S2 route, best route)
(destroy house, B3 and B4 Corridor, not want)
......
(lose the home, Illiana Corridor, absurd)
......
Final Mobility and Accessibility Cluster
(move truck off local route, Illiana, NA)
(commuter, entire section of 281, nightmare)
......
(increase access to job, build alternative, support)

Figure 7.3 – Examples of Opinion Tuples Clustering for Concern Classification

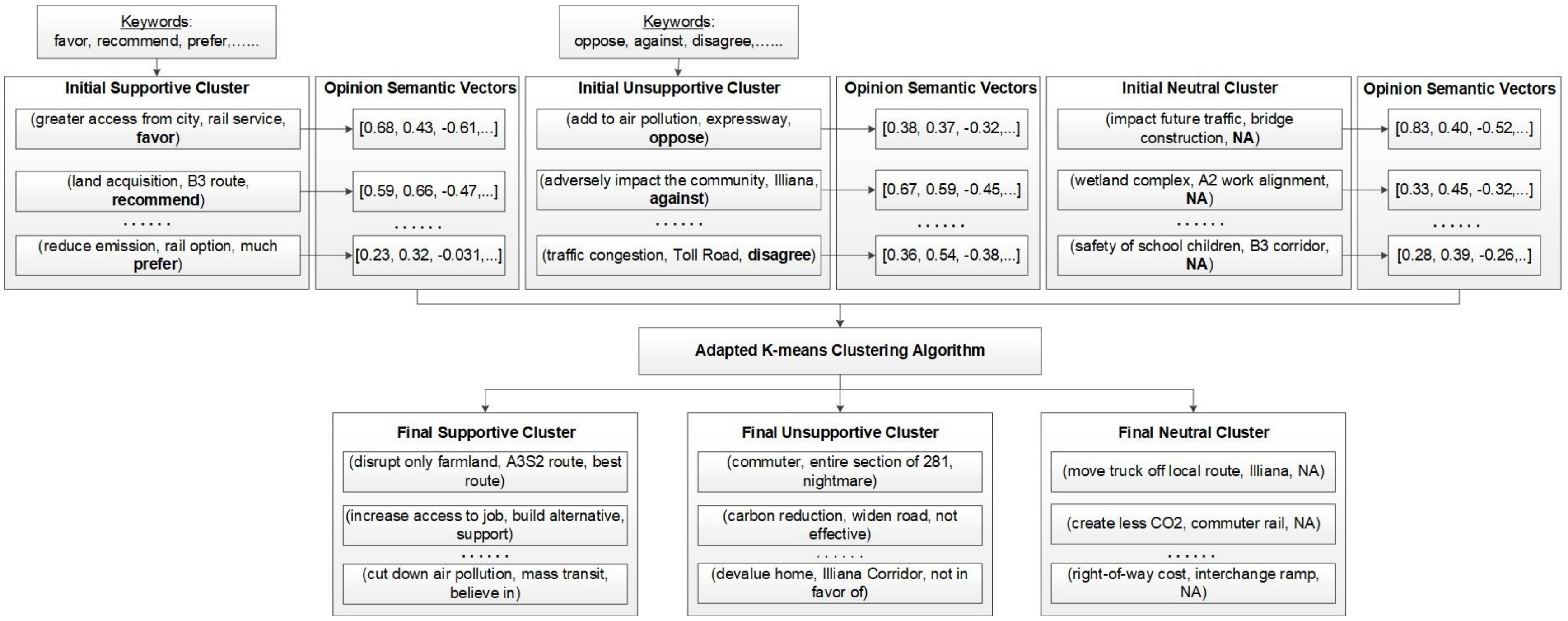

Keywords:
favor, recommend, prefer,……
Keywords:
oppose, against, disagree,……
Initial Supportive Cluster
(greater access from city, rail service, favor)
(land acquisition, B3 route, recommend)
……
(reduce emission, rail option, much prefer)
Opinion Semantic Vectors
[0.68, 0.43, -0.61,...]
[0.59, 0.66, -0.47,...]
……
[0.23, 0.32, -0.031,...]
Initial Unsupportive Cluster
(add to air pollution, expressway, oppose)
(adversely impact the community, Illiana, against)
……
(traffic congestion, Toll Road, disagree)
Opinion Semantic Vectors
[0.38, 0.37, -0.32,...]
[0.67, 0.59, -0.45,...]
……
[0.36, 0.54, -0.38,...]
Initial Neutral Cluster
(impact future traffic, bridge construction, NA)
(wetland complex, A2 work alignment, NA)
……
(safety of school children, B3 corridor, NA)
Opinion Semantic Vectors
[0.83, 0.40, -0.52,...]
[0.33, 0.45, -0.32,...]
……
[0.28, 0.39, -0.26,..]
Adapted K-means Clustering Algorithm
Final Supportive Cluster
(disrupt only farmland, A3S2 route, best route)
(increase access to job, build alternative, support)
……
(cut down air pollution, mass transit, believe in)
Final Unsupportive Cluster
(commuter, entire section of 281, nightmare)
(carbon reduction, widen road, not effective)
……
(devalue home, Illiana Corridor, not in favor of)
Final Neutral Cluster
(move truck off local route, Illiana, NA)
(create less CO2, commuter rail, NA)
……
(right-of-way cost, interchange ramp, NA)

Figure 7.4 – Examples of Opinion Tuples Clustering for Sentiment Classification

The initialization step prepares the initial members of each classification category. For each category, the opinion tuple that contains at least one corresponding keyword is identified as relevant to the category, where all relevant opinion tuples become initial members of the category's cluster. For concern classification, all opinion tuples without concern expressions are initially assigned to the "general concern" cluster. For sentiment classification, all opinion tuples without sentiment expressions are initially assigned to the "neutral" cluster. If an opinion tuple contains a keyword and a negation word (e.g., not, no, never) in its opinion expression, the opinion tuple is assigned to the opposite sentiment category of the keyword. For example, if an opinion tuple has an opinion expression of "not in favor of", which contains a keyword of the supportive category ("favor") and a negation word ("not"), then the opinion tuple is assigned to the unsupportive category. A list of negation words were compiled and utilized, including "not", "no", "never", "neither", "nor", "none", "no one", "nobody", "hardly", and "rather".

After the initialization step, the iteration step is conducted to assign each opinion tuple in the training data to the existing cluster(s) and update the clusters accordingly, in an iterative manner. At each iteration, an opinion tuple *O_i* in the training data is compared and assigned with existing clusters based on its semantic similarity with each cluster. As defined in Eq. (7.5), the semantic similarity between an opinion tuple O_i and an opinion tuple cluster C_k is denoted as $Sim(O_i, C_k)$, and is calculated as the cosine similarity between the opinion semantic vector V_i for opinion tuple O_i and $\bar{C}_k$, which is the centroid of the cluster C_k. $\bar{C}_k$ is defined in Eq. (7.6), where V_x is the opinion semantic vector for an opinion tuple O_x in the cluster C_k, and y is the total number of opinion tuples in the cluster C_k.

$$Sim(O_i, C_k) = \frac{V_i * \bar{C}_k}{\|V_i\| * \|\bar{C}_k\|} \tag{7.5}$$

$$C_k = \{\frac{\sum_{x=1}^{y} V_x}{y}\} \quad (7.6)$$

For concern classification, because an opinion tuple O_i can have more than one concern label, a threshold value T_k is introduced to determine whether to assign O_i to a cluster C_k. If the $Sim(O_i, C_k)$ is greater than or equal to the threshold value T_k, the opinion tuple O_i is assigned to the cluster C_k. For a cluster C_k, the threshold T_k is the mean of the semantic similarities between the cluster C_k and other clusters. T_k is defined in Eq. (7.7), where $Sim(C_k, C_l)$ is the semantic similarity between the cluster C_k and another cluster C_l, and n is the total number of clusters. As defined in Eq. (7.8), $Sim(C_k, C_l)$ is calculated as the cosine similarity between $\bar{C_k}$ and $\bar{C_l}$, which are the centroids of the cluster C_k and a cluster C_l, respectively. If no existing cluster has a semantic similarity over the corresponding threshold value, the opinion tuple O_i is assigned to the cluster with the largest semantic similarity.

$$\mathrm{T}_k = \frac{1}{n-1} \sum_{l=1, l \neq k}^{n} Sim(C_k, C_l) \quad (7.7)$$

$$Sim(C_k, C_l) = \frac{\bar{C_k} * \bar{C_l}}{\|\bar{C_k}\| * \|\bar{C_l}\|} \quad (7.8)$$

Because an opinion tuple O_i can only have one sentiment label, for sentiment classification, the opinion tuple O_i would be assigned to the cluster with the largest semantic similarity to the opinion tuple O_i.

After assigning the opinion tuple O_i to one or more clusters, the centroid(s) of the new cluster(s) needs to be updated before proceeding to the next opinion tuple. The iteration step stops when the centroid of each cluster stabilizes and there is no change in the assignment of each opinion tuple.

7.2.4 Opinion Classification

The generated opinion tuple clusters serve as labeled training data for opinion classification using a supervised machine learning algorithm. Because the quality of training data can have a significant impact on the performance of the supervised algorithm, the semantic similarity between an opinion tuple and the cluster(s) it belongs was selected as a criterion to determine whether the opinion tuple should be included as part of the labeled training data. When the final clusters are formed, the opinion tuples of each concern category are ranked in descending order based on their semantic similarities with the corresponding cluster. Only the opinion tuples with the top p semantic similarity become part of the labeled training data. To evaluate the impact of p on the classification performance, p values from 50% to 100% with a 10% interval were tested.

In this research, the concern classification is a multilabel text classification problem, where each opinion tuple can be classified into one or more concern categories. Two supervised machine learning algorithms were thus tested: the SVM algorithm (as a representative of the PTM) and BP-MLL algorithm (as representative of the AAM). For the SVM, to avoid the data imbalance problem, a multiclass approach was adopted to each subproblem instead of a binary classification approach. The multiclass approach requires a multiclass classifier for each label, where a subset of the training data was used. For each subproblem with label L_k, the training dataset includes all the opinion tuples assigned with the label L_k and other opinion tuples that are assigned with a single label other than L_k. Because sentiment classification is a multiclass text classification problem, where each opinion tuple can only be labeled as “supportive”, “unsupportive” or “neutral”, the multiclass approach was also adopted.

7.2.5 Experimental Setup

7.2.5.1 Data Preparation

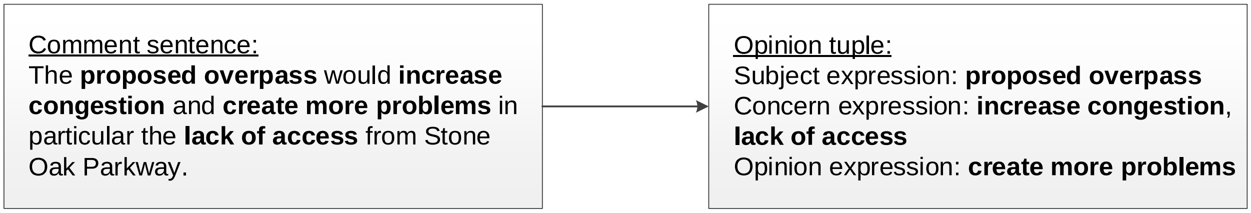
To ensure that the comment collection covers a variety of stakeholder concerns, nine large-scale transportation projects from eight states were selected (see Table 7.2). For these projects, the comments that were received during the public comment period, including comments submitted through project websites, public hearings, emails, and social media, were gathered into a comment collection. The comment collection contains 3,132 comments in total. A total of 520 comments were randomly selected – 400 for training and 120 for testing – from the collection, which include 1,823 and 460 sentences, respectively.

Table 7.2 – Statistics about the Comment Data Collection

Project name	Project location	Number of comments	Number of sentences	Average sentences per comment
Cleveland Opportunity Corridor	Ohio	140	399	3
I-395 Transportation System	Maine	136	406	3
Illiana Corridor Tier 1	Illinois & Indiana	1,129	8,569	8
OR62 Corridor	Oregon	64	407	6
US281	Texas	641	5,725	9
Crosstown Parkway	Florida	37	336	9
Gulf Coast Parkway	Florida	43	346	8
I-5	California	343	2,100	6
North I-25	Colorado	599	3,958	7
Total	NA	3,132	22,246	7

To prepare the gold standard, the opinion tuples in the training and testing datasets were manually extracted and classified. An opinion tuple includes three parts: subject expression, concern expression, and opinion expression parts. For concern classification, each opinion tuple was manually classified into one or more stakeholder concern categories. For sentiment classification, each opinion tuple was manually classified into one and only one sentiment category. For example, a comment sentence and its corresponding opinion tuple and their classifications are shown in

Figure 7.5. In this example, the opinion tuple has a subject expression "proposed overpass", two concern expressions "increase congestion" and "lack of access", and an opinion expression "create more problem", which were assigned to the "mobility and accessibility" concern category and the "unsupportive" sentiment category.



Comment sentence:
The **proposed overpass** would **increase congestion** and **create more problems** in particular the **lack of access** from Stone Oak Parkway.

Opinion tuple:
Subject expression: **proposed overpass**
Concern expression: **increase congestion**, **lack of access**
Opinion expression: **create more problems**

Figure 7.5 – An Example of a Comment Sentence, its Opinion Tuples, and their Classifications

The gold standard labels for the opinion tuples were determined based on mutual agreement among three annotators (the author and another two researchers). The number of opinion tuples for each concern and sentiment category are shown in Table 7.3.

Table 7.3 – Number of Comments for each Concern and Sentiment Category

Category	Number of opinion tuples
Concern category	
Air quality	102
Water resource	79
Wildlife and habitat	98
Noise control	175
Traffic	168
Mobility and accessibility	165
Physical infrastructure	310
Transportation safety	246
Cost and funding	286
Land use and property	175
Regional development	119
Cultural concern	129
Management/administrative concern	174
General concern	201
Sentiment category	
Supportive	653
Unsupportive	897
Neutral	339

7.2.5.2 Data Preprocessing

Two data preprocessing techniques were utilized: tokenization and lemmatization. Tokenization breaks the opinion tuple into meaningful tokens (terms). Lemmatization transforms a word into its base or dictionary form (i.e., lemma). For example, after the lemmatization, the words “preferring”, “preferred”, and “prefers” would all be transformed into their lemma “prefer”. Lemmatization can reduce the number of features by grouping the words with the same lemma, and can in turn be effective when generating opinion tuple clusters.

7.2.5.3 Supervised Machine Learning Algorithm Implementation

For opinion classification, six supervised machine algorithms were implemented and tested: SVM, NB, RF, ME, ML-KNN, and BP-MLL. The term frequency and inverse document frequency (TF-IDF) was used for term weighting, because it is the state-of-the-art weighting scheme for TC. Term weighting assigns numerical values to the terms in a document, which represent how much these terms contribute to the semantics of the document (Lan et al. 2009). The SVM algorithm with linear kernel was adopted, because previous studies (Hsu et al. 2003) indicate that linear kernels tend to perform well and nonlinear kernels do not necessarily offer significant performance improvement when solving problems with a large number of features, such as text classification. For each algorithm, parameter tuning was conducted to optimize the classification performance empirically. For example, the ML-KNN algorithm has two important parameters: *K*, which indicates the number of nearest neighbors used for classifying an unseen instance, and a smoothing factor *S*, which controls the strength of the uniform prior in determining the posterior probability. To optimize the parameters *K* and *S*, a range of values for each parameter (e.g., 1 to 5 with an interval of 1 for *K*, and 0 to 1 with an interval of 0.1 for *S*) was evaluated and the best combination of values (k=2 and S=0.5) was selected based on the classification performance (F1 measure). The

SVM, NB, RF, and ME algorithms were implemented using the scikit-learn ML in Python package (Pedregosa et al. 2011); the ML-KNN was implemented using scikit-multilearn multilabel classification for Python package (Szymanski and Kajdanowicz 2017); and the BP-MLL algorithm was implemented using the MATLAB package for multilabel BP neural networks (Zhang and Zhou 2006).

7.2.6 Evaluation

For concern classification, the example-based multilabel evaluation metrics were adopted. Example-based precision and recall were calculated using Eq. (7.9) and Eq. (7.10), where TP_i is the number of labels assigned correctly as positive for opinion tuple O_i; FP_i is the number of labels assigned incorrectly for opinion tuple O_i; FN_i is the number of labels assigned incorrectly as negative for opinion tuple O_i; and t is the total number of testing opinion tuples.

$$Example - based\ Precision = \frac{1}{t}\sum_{i=1}^{t}\frac{TP_i}{TP_i+FP_i} \tag{7.9}$$

$$Example - based\ Recall = \frac{1}{t}\sum_{i=1}^{t}\frac{TP_i}{TP_i+FN_i} \tag{7.10}$$

For sentiment classification, both the supervised and unsupervised methods were evaluated using precision, recall, and F1 measure, as per Eqs. (7.11), (7.12), and (7.13), where true positive (TP) refers to the number of opinion tuples classified correctly, false positive (FP) refers to the number of opinion tuples classified incorrectly, and false negative (FN) refers to the number of opinion tuples incorrectly classified as negative. Precision, here, is defined as the ratio of the number of correctly classified opinion tuples over the total number of classified opinion tuples. Recall, here, is defined as the ratio of the number of correctly classified opinion tuples over the total number of opinion tuples that should be classified. F1 is the harmonic mean of precision and recall. These measures were calculated based on a comparison of the experimental results with the manually-developed gold standard.

$$Precision = \frac{TP}{TP+FP} \tag{7.11}$$

$$Recall = \frac{TP}{TP+FN} \tag{7.12}$$

$$F1\ measure = \frac{2*Precision*Recall}{Precision+Recall} \tag{7.13}$$

7.3 Experimental Results and Analysis

7.3.1 Opinion Semantic Vectors: Selecting the Word-Embedding Model and Corpus

For the proposed unsupervised method, when trained on the top 50% of the clustered opinion tuples, the performance of concern and sentiment classification utilizing the four different opinion semantic vectors with the supervised machine learning algorithms are summarized in Table 7.4. As shown in Table 7.4, the performance when utilizing the opinion semantic vectors learned from the stakeholder comment collection is generally better than those learned from Wikipedia. This is because Wikipedia is not a domain-specific corpus, and may not be able to capture the semantic information and relationships between terms in the infrastructure domain well. For example, when utilizing the vectors learned from the comment collection, both the skip-gram and GloVe models showed greater semantic similarity between the concern expressions "acquisition of land" and "highway right-of-way". In terms of the word-embedding models, the skip-gram model outperformed the GloVe model on F1 measure when utilizing the same corpus and the same classifier. This is likely because the skip-gram model creates word embeddings based on the context information of terms, and thus has a better capability to capture semantic similarities between opinion tuples than the GloVe model, which mainly utilizes term cooccurrence statistics.

7.3.2 Opinion Classification: Selecting the Supervised Machine Learning Algorithm

In terms of the supervised machine learning algorithm for concern classification, on average, the SVM algorithm achieved 82%, 85%, and 83% example-based precision, recall, and F1 measure, respectively, and outperformed the average performance of the BP-MLL algorithm (81%, 83%, and 81% example-based precision, recall, and F1 measure, respectively). This is likely arising from two reasons. First, the SVM algorithm adopted a multiclass classification approach, which avoids the data imbalance problem. Second, the BP-MLL algorithm generally works better when the training data show strong label correlations, which is not the case for the generated opinion tuple clusters. For example, only 32% of the opinion tuples have more than one label, and the average Pearson correlation coefficient for each label pair is approximately -0.02, which indicates very low correlations between label pairs.

Overall, when trained on the top 50% of the clustered opinion tuples, the best performance was achieved when using the skip-gram word-embedding model for learning the opinion semantic vector representations of words from the comment collection and utilizing the SVM algorithm for classification (PTM with multiclass SVM for concern classification and multiclass SVM for sentiment classification). These performance results are 85%, 88%, and 87% example-based precision, recall, and F1 measure for concern classification; and 84%, 83%, and 84% precision, recall, and F1 measure for sentiment classification.

Table 7.4 – Performance of the Proposed Unsupervised Method on Classification

Proposed unsupervised method (trained on top 50% of the clustered opinion tuples)			Precision	Recall	F1 measure
Word semantic vector model	Corpus	Supervised ML algorithm			
Concern classification					
Skip-gram	Wikipedia	SVM	80%	81%	80%
GloVe	Wikipedia		79%	81%	80%
Skip-gram	Collection*		85%	88%	87%
GloVe	Collection*		84%	88%	86%
		Average	82%	85%	83%
Skip-gram	Wikipedia	BP-MLL	78%	78%	78%
GloVe	Wikipedia		77%	78%	77%
Skip-gram	Collection*		85%	87%	86%
GloVe	Collection*		82%	87%	84%
		Average	81%	83%	81%
Sentiment classification					
Skip-gram	Wikipedia	Multiclass SVM	78%	77%	78%
GloVe	Wikipedia		79%	79%	79%
Skip-gram	Collection*		84%	83%	84%
GloVe	Collection*		83%	83%	83%
Average			81%	81%	81%

* Highway stakeholder comments collection

7.3.3 Opinion Tuple Clustering: Selecting the Portion of Clusters to Use for Training

Using the aforementioned combination (skip-gram, comment collection, and SVM), the impact of varying the percentage of clustered opinion tuples used for training was further evaluated. The results are summarized in Table 7.5. Values from 50% to 100%, with a 10% interval, were tested.

The best concern classification performance – 88%, 90%, and 89% exampled-based precision, recall, and F1 measure, respectively – was achieved using the top 80% of the tuples for training. From 50% to 80%, the performance gradually improved with the increase in percentage, up to the optimal point of 80%, after which the performance started to decline. This is because the top 50% to 80% tuples contain more effective examples (opinion tuples that are assigned to the correct cluster) than noisy examples (opinion tuples that are assigned to the incorrect cluster, which typically have relatively smaller semantic similarities with the incorrect cluster); while the top 80% to 100% tuples often contain more noisy examples than effective examples.

The best sentiment classification performance – 87%, 86%, and 86% precision, recall, and F1 measure, respectively – was achieved using the top 70% of the tuples for training. Compared with concern classification, the optimal point for sentiment classification is lower (70% vs. 80%), which indicates that the sentiment clusters contain more noisy examples compared with the concern clusters.

Table 7.5 – Effect of Percentage of Opinion Tuples Used for Training on Classification Performance

Percentage	Precision	Recall	F1 measure	Standard deviation of F1 measure
Concern classification				
50%	85%	88%	87%	4.2%
60%	87%	89%	88%	4.4%
70%	87%	89%	88%	4.5%
80%	88%	90%	89%	4.7%
90%	86%	87%	86%	4.7%
100%	84%	86%	85%	4.2%
Sentiment classification				
50%	84%	83%	84%	3.4%
60%	85%	85%	85%	3.2%
70%	87%	86%	86%	4.1%
80%	85%	85%	85%	3.7%
90%	85%	84%	85%	3.2%
100%	84%	84%	84%	3.5%

7.3.4 Overall Performance of the Proposed Unsupervised Method and Comparison with Existing Methods

Based on the aforediscussed experimental results, the proposed unsupervised method for opinion classification (1) uses the skip-gram word-embedding model for learning opinion semantic vector representations of words from a stakeholder comment collection, (2) creates opinion tuple clusters for each category based on these learned vectors, and (3) trains the SVM algorithm on the top *p*% – 80% for concern classification and 70% for sentiment classification – of the clustered tuples.

The proposed unsupervised method was then compared with both the PTM and AAM for multilabel concern classification. Four popular algorithms for PTM were selected, including NB, RF, ME, and SVM. All the algorithms implemented for the PTM adopted the same multiclass approach as the proposed unsupervised method. Two popular algorithms for AAM were also selected for comparison, including ML-KNN and BP-MLL. The performance results are summarized in Table 7.6. The proposed unsupervised method achieved the best example-based recall, and the second-best example-based precision and F1 measure – second by a small margin only (88% vs. 90% for precision and 89% vs. 90% for F1 measure). These results show that the proposed unsupervised method achieved a similar level of performance – even better performance for recall – to that of the best-performing supervised method. This means that the proposed method can offer a similar (or even improved) level of performance, while saving a lot of manual effort in labeling.

The proposed unsupervised method was also compared with the supervised method for multiclass sentiment classification. Four popular multiclass classification algorithms were selected, including multiclass NB, RF, ME, and SVM. The performance results are summarized in Table 7.6. The proposed unsupervised method achieved the second-best performance in terms of precision, recall, and F1 measure. These results show that the proposed unsupervised method also achieved a comparable level of performance (86% vs. 89% for F1 measure) for sentiment classification – compared to that of the best-performing supervised method. Compared to concern classification, however, sentiment classification showed a lower level of performance (86% vs. 89% for F1 measure) because the quality of the sentiment clusters depends largely on the choice of keywords.

Compared with concern expressions, the sentiment expressions that represent a sentiment category can take many different forms and can use a wider variety of terms. Therefore, the use of positive and negative opinion words from a sentiment lexicon as keywords may not represent the full spectrum of a sentiment category. For example, opinion tuples that contain uncommon sentiment expressions such as “first choice”, “draws blanket conclusion”, and “boondoggles” were mislabeled into other categories.

Table 7.6 – Comparison between the Proposed Unsupervised Method and Existing Supervised Methods based on Classification Performance

Opinion classification method	Precision	Recall	F1 measure
Concern classification			
NB*	81%	84%	82%
RF*	78%	85%	81%
ME*	81%	84%	82%
SVM*	90%	89%	90%
ML-KNN**	83%	84%	83%
BP-MLL**	87%	88%	87%
Proposed unsupervised method	88%	90%	89%
Sentiment classification			
Multiclass NB	82%	82%	82%
Multiclass RF	82%	80%	81%
Multiclass ME	85%	84%	84%
Multiclass SVM	89%	88%	89%
Proposed unsupervised method	87%	86%	86%

* Problem transformation methods

** Algorithm adaption methods

7.3.5 Error Analysis

Two main types of errors were identified based on the testing results. First, opinion tuples with implicit concerns were misclassified. For example, an opinion tuple has a subject expression “A3S2 corridor”, and a concern expression “residential and business displacement”. It should be classified into the “land use and property” category, because it expresses concerns over businesses and residents whose properties would be displaced due to the project land use. However, because the opinion tuple does not explicitly mention any words related to this category (e.g., “land”, “home”, or “property”), it was mistakenly classified into the “general concern” category. To address this type of error, some strategies could be considered and tested in future work. For example, more opinion tuples with implicit concerns could be included in the training data and/or the concern key words could be expanded to include more implicit terms/phrases. Second, opinion tuples that have uncommon opinion expression were misclassified. For example, an opinion tuple has a subject expression “toll road”, a concern expression “congestion”, and an opinion expression “boondoggles”. It expresses unsupportive sentiment towards the toll road, but was mistakenly classified into neutral category because “boondoggles” is not a negative opinion word in the sentiment lexicon. In future work, the sentiment lexicon could be expanded to include more uncommon opinion words/phrases.

CHAPTER 8: SENTENCE-LEVEL STAKEHOLDER OPINION MINING

8.1 State of the Art and Knowledge Gaps

Stakeholder opinion mining is the process of discovering patterns or knowledge from unstructured stakeholder opinions (Montoyo et al. 2012; Ravi and Ravi 2015). A stakeholder opinion is a piece of text that expresses the attitude of a stakeholder towards a target object, such as a movie, a restaurant, or a highway project, in this research. Sentence-level stakeholder opinion mining has long been studied in the computer science domain. For example, Khan et al. (2013) developed a Twitter opinion mining framework that involves various preprocessing techniques and a hybrid scheme of classification algorithms. Sayeedunnissa et al. (2013) developed a Boolean classification model for conducting opinion mining on social network using a naïve Bayes algorithm and bag-of-word features. Yang and Cardie (2014) proposed a context-aware method for analyzing sentiment through modeling of complex linguistic structures and capturing both local and global contextual information. To conduct sentiment polarity categorization on Amazon reviews at both sentence and document level, Fang and Zhan (2015) proposed a negation phrase identification algorithm, a sentiment score computation method, and a feature vector generation method. Appel et al. (2016) used a hybrid approach to sentence-level sentiment analysis that utilizes a sentiment lexicon enhanced with SentiWordnet, a set of semantic rules, and fuzzy sets to estimate the semantic orientation polarity and its intensity.

Because sufficient labeled training data can be hard or costly to obtain, many efforts have taken an unsupervised approach. For example, Hu et al. (2013) proposed an unsupervised sentiment analysis framework that incorporated emotion signals as prior knowledge to guide the learning process through modelling word-level and post-level emotion indication and correlation. Marrese-Taylor et al. (2014) developed an unsupervised tourism opinion mining system based on the extension of Liu’s (2007) aspect-based opinion mining method, and proposed complex natural language processing (NLP) rules to account for the linguistic features of tourism product reviews when determining the sentiment of identified aspects. Jimenez-Zafra et al. (2015) proposed an unsupervised approach for aspect-level sentiment analysis that determines the sentiment expressed on an aspect based on the polarity of each modifier word calculated through a voting of three sentiment classifiers. Garcia-Pablos et al. (2017) developed an almost unsupervised system for multi-domain and multi-lingual aspect-level sentiment analysis, which combines the topic-modelling with the continuous word embeddings and a Maximum Entropy classifier, and requires a minimal set of seed words per target domain and language.

Outside of the computer science domain, a number of research studies have been conducted on applying opinion/text mining techniques in the construction domain. Choudhary et al. (2009) applied text mining techniques (e.g., feature extraction, information retrieval, and text categorization) to uncover patterns, associations, and trends from post-project reviews (PPRs) collected by construction companies. Ur-Rahman and Harding (2011) proposed a text mining system that combines clustering techniques and a priori association rule mining to improve the classification of PPRs collected from the construction industry. Fan and Li (2012) utilized text mining techniques to represent unstructured textual cases by a structured vector model to retrieve relevant historical construction accident cases from a case library. Nik-Bakht and El-Diraby (2016) utilized community detection techniques with information retrieval methods to analyze the followers of the Toronto Light Rail Transit project on Twitter, profile them based on their interests and opinions, and monitor the dynamics of the follower communities and opinions.

Despite the abovementioned research efforts, three main knowledge gaps are identified. First, existing sentence-level opinion mining methods mostly focus on sentiment analysis, and have limited ability to identify stakeholder concerns from their comments. Second, existing efforts that took an unsupervised approach mostly rely on NLP rules or algorithms such as topic-modelling, which are not suitable for classifying stakeholder opinions into precise and fine-grained concern categories to support highway decision making. Third, there is lack of research efforts that applied stakeholder opinion mining in the highway infrastructure domain to discover new patterns and knowledge about the opinions of stakeholders on real-life infrastructure projects, and how these opinions differ from one stakeholder group to another.

8.2 Proposed Sentence-Level Stakeholder Opinion Mining Method

To address the aforementioned knowledge gaps, this research proposes an unsupervised machine learning-based sentence-level stakeholder opinion mining method, which can automatically create pseudo training data through topic model-based concern labeling and lexicon-based sentiment labeling. Compared to the tuple-based method (Chapters 6 and 7), the sentence-level method offers an alternative approach when a sentence-level analysis is sufficient. A supervised classifier then learns from the automatically-created pseudo training data to classify comment sentences into one or more concern categories and one sentiment category. The proposed method includes three primary elements: concern labeling, sentiment labeling, and supervised opinion classification.

8.2.1 Concern Labeling

The concern labeling aims to assign tentative concern labels to each comment sentence based on their respective concern confidence scores, which indicate the probabilities of a comment sentence expressing specific stakeholder concerns. The concern confidence scores of the comment sentences are estimated based on their respective concern-topic distributions. The LDA model (Blei et al. 2003) and the collapse Gibbs sampling method (Griffiths and Steyvers 2004) were adapted to learn the concern-topic distributions. They were adapted in the following ways: (1) using pre-defined seed words to guide the topic assignment at the initialization stage, and (2) integrating semantic similarities into the topic-word distribution update at the iteration stage.

So, at initialization stage, each word in every comment sentence is assigned to one of the *K* concern topics. If the word is a seed word of a topic, it gets assigned to that topic. If not, it gets randomly assigned to one of the *K* topics. The seed words are pre-defined words that represent each concern topic, and are used to provide guidance for the concern labeling process. Seed words can be defined based on existing domain knowledge in the form of keyword lists, thesauri, taxonomies, ontologies, etc. For this research, the seed words were empirically defined based on the stakeholder concern hierarchy. They were defined based on the terms in the names of the concern concepts and subconcepts and the synonyms of these terms. Fourteen (14) stakeholder concern categories were used for concern labelling based on the stakeholder concern hierarchy (Figure 8.1). As mentioned in Chapter 4, the hierarchy was developed based on a literature review on transportation decision making and stakeholder involvement processes, and interactions with transportation practitioners during one-to-one interviews. Table 8.1 shows the seed words of the four concern categories.

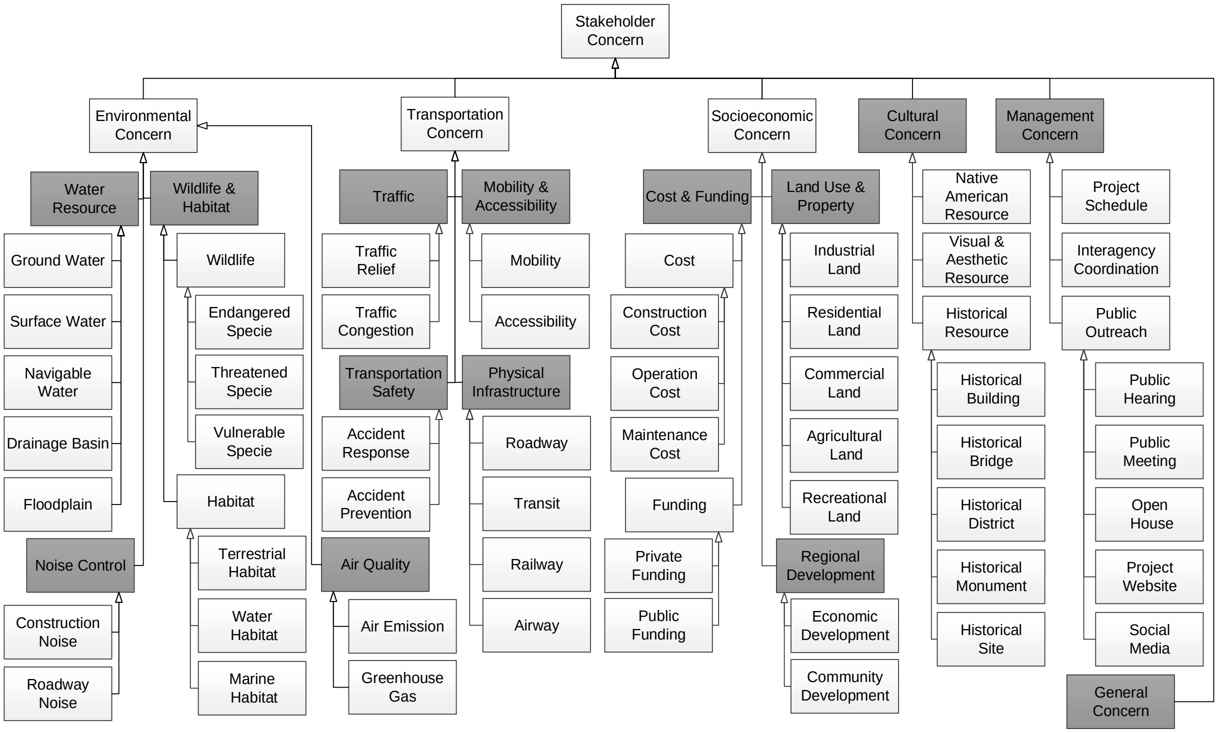


Figure 8.1 – A Partial View of the Stakeholder Concern Hierarchy

Table 8.1 – Sample Concern Category Keywords

Concern category	Keywords
Air quality	Air, emission, CO2, PM2.5, PM10, atmosphere, ozone, greenhouse
Water resource	Water, wetland, river, stream, creek, lake, marsh, floodplain, lagoon
Wildlife and habitat	Wildlife, habitat, specie, animal
Physical infrastructure	Infrastructure, highway, tollway, freeway, bridge, railway, overpass, station, airport, terminal

At the iteration step, the concern topic assigned to each word in the comment sentence is updated based on the learned topic-document and topic-word distributions. To account for the semantic similarities between different words and concern topics, a word-topic parameter (a parameter that varies for each word-topic pair) is proposed and integrated with the topic-word distribution update. For a word w_i and a concern topic k, the topic-word distribution $\emptyset_k^{w_i}$ is modified and defined in Eq. (8.1), where $n_{-i,k}^{w_i}$ is the number of times w_i is assigned to topic k in the comment collection excluding the current w_i, $\sum_{i=1}^{W} n_{-i,k}^{w_i}$ is the total number of words that are assigned to topic k in the comment collection excluding the current w_i, W is the total number of words in the comment collection, and $\beta_{i,k}$ is a word-topic parameter that controls the topic distribution based on the semantic similarities between w_i and the seed words of topic k.

$$\emptyset_k^{w_i} = \frac{n_{-i,k}^{w_i} + \beta_{i,k}}{\sum_{i=1}^{W} n_{-i,k}^{w_i} + W\beta_{i,k}} \tag{8.1}$$

The word-topic parameter, $\beta_{i,k}$, is proposed and defined in Eq. (8.2) based on the notion that words similar to the seed words of a topic should be more likely to be assigned to that topic. As per Eq. (8.2), w_l^k is a seed word for topic k, $Sim(w_i, w_l^k)$ is the semantic similarity between w_i and w_l^k, Q is the total number of seed words for topic k, and β is a value between 0 and 1 that sets the upper bound of $\beta_{i,k}$ and is experimentally determined.

$$\beta_{i,k} = \frac{\sum_l^Q Sim(w_i, w_l^k)}{Q} \beta \tag{8.2}$$

As defined in Eq. (8.3), $Sim(w_i, w_l^k)$ is calculated as the cosine similarity between $\overline{w}_i$ and $\overline{w}_i^k$, which are the word embeddings of w_i and w_l^k, respectively.

$$Sim\left(w_i, w_l^k\right) = \frac{\overline{w}_i * \overline{w}_i^k}{\|\overline{w}_i\| * \|\overline{w}_i^k\|} \tag{8.3}$$

The word embedding of a word is a real-valued vector of features that characterize the meaning and context information in which the word appears in a corpus of text. In this research, the skip-gram model (Mikolov et al. 2013) was used to develop the word embeddings based on the stakeholder comment collection.

The iteration stops when the reassignment probability of each word stabilizes. For a sentence $d_j$and a topic k, the concern confidence score $Con(c_{d_j} = k, d_j)$ is defined in Eq. (8.4) (Griffiths and

Steyvers 2004), where $n_{d_j}^k$ is the number of words assigned to topic k in d_j, $\sum_{k=1}^{K} n_{d_j}^k$ is the sum of the number of words assigned to each topic in d_j, K is the total number of topics, and α is a parameter that controls the topic-document distribution.

$$Con(c_{d_j} = k, d_j) = \frac{n_{d_j}^k + \alpha}{\sum_{k=1}^{K} n_{d_j}^k + K\alpha} \tag{8.4}$$

For a sentence d_j and a topic k, if $Con(c_{d_j} = k, d_j)$ is greater than or equal to the threshold T_k, d_j is labeled with concern k. If none of the concern confidence score surpasses the threshold, d_j is labeled with "general concern". The threshold T_k is defined in Eq. (8.5), where $Con(c_{d_j} = l, d_j)$ is the concern confidence score between d_j and another topic l, and K is the total number of concern topics.

$$\mathrm{T}_k = \frac{1}{K-1} \sum_{l=1, l \neq k,}^{K} Con(c_{d_j} = l, d_j) \tag{8.5}$$

8.2.2 Sentiment Labeling

Sentiment labeling aims to assign sentiment labels (supportive, neutral, and unsupportive) to each comment sentence based on their respective sentiment confidence scores. A lexicon-based method is proposed to integrate word-level sentiments and word-negation relations to estimate the sentiment confidence scores at the sentence level.

For each comment sentence, the sentiment confidence scores are calculated based on the Sentiwordnet 3.0 (Baccianella et al. 2010), which assigns positive and negative real-valued sentiment scores to each word that belongs to a WordNet synset. For a comment sentence d_j, the supportive sentiment score $S\ (d_j)^+$ and unsupportive sentiment score $S\ (d_j)^-$are defined as per Eqs. (8.6) and (8.7), where $S(w_i)^+$ is the positive sentiment score for the word w_i, $S(w_i)^-$is the negative sentiment score for w_i, m_{w_i} is the negation modifier for w_i, and q is the total number of words in d_j. For a word w_i, both its positive and negative sentiment scores are obtained from the Sentiwordnet 3.0 (Baccianella et al. 2010). If the word does not belong to a WordNet synset, it gets zero positive and negative sentiment scores. The negation modifier m_{w_i} indicates whether the sentiment orientation of the word w_i is modified by a negation word, such as "no", "not", and "nobody". If modified, the value of the m_{w_i} equals to 1, otherwise it equals to 0. In this research, two common negation contexts are considered: direct negation and indirect (long-distance) negation. The direct negation refers to the scenario where a word or the phrase that contains the word directly follows a negation word. For example, in the comment sentence "I do not support any plan (now or future) for any form of toll road", the word "support" is directly negated by "not". Indirect negation refers to the scenario where the directly negated word is followed by a complement clause. For example, in the comment sentence "I do not think this project would be a success", the word "think" is directly-negated by "not". Because the complement clause "this project would be a success" follows the directly-negated word "think", the subject word of the clause "success" would be indirectly negated. The Stanford dependency parser (Manning et al. 2014) was utilized to identify the negated words, in the direct and indirect negation contexts.

$$S\ (d_j)^+ = \sum_i^q (S(w_i)^+ * (1 - m_{w_i}) + S(w_i)^- * m_{w_i}) \tag{8.6}$$

$$S\ (d_j)^- = \sum_i^q (S(w_i)^- * (1 - m_{w_i}) + S(w_i)^+ * m_{w_i}) \tag{8.7}$$

The supportive and unsupportive sentiment confidence scores $Sent(d_j)^+$ and $Sent(d_j)^-$ are defined in Eqs. (8.8) and (8.9), where $S\ (d_j)^+$ and $S\ (d_j)^-$ are the supportive and unsupportive sentiment scores for sentence d_j, respectively. $Sent(d_j)^+$and $Sent(d_j)^-$represent the likelihood that d_j expresses supportive and unsupportive sentiments, respectively. If the difference between

$Sent(d_j)^+$ and $Sent(d_j)^-$ is greater than the threshold value T, then d_j is labeled with the supportive sentiment. If the difference between $Sent(d_j)^-$ and $Sent(d_j)^+$ is greater than T, then d_j is labeled with the unsupportive sentiment. Otherwise, d_j is labeled with the neutral sentiment.

$$Sent(d_j)^+ = \frac{e^{S\,(d_j)^+}}{e^{S\,(d_j)^+} + e^{S\,(d_j)^-}} \tag{8.8}$$

$$Sent(d_j)^- = \frac{e^{S\,(d_j)^-}}{e^{S\,(d_j)^+} + e^{S\,(d_j)^-}} \tag{8.9}$$

8.2.3 Supervised Opinion Classification

In the proposed method, the labeled comment sentences serve as pseudo training data for opinion classification using a supervised machine learning algorithm. Because the unsupervised labeling process – for both concern and sentiment labeling – could produce noisy data (comment sentences with incorrect labels), which can undermine the performance of the machine learning algorithm, only a subset of the created pseudo training data are used. The subset is selected based on the concern and sentiment confidence scores. For each concern and sentiment category (except for the neutral category), the comment sentences labeled with the category are ranked in descending order based on their respective concern/sentiment confidence scores. The comment sentences labeled with the neutral category are ranked in ascending order based on the absolute value of the difference between their supportive and unsupportive sentiment confidence scores. Only the comment sentences with the top *p* confidence scores become part of the pseudo training data. To evaluate the impact of *p* on the classification performance, *p* values from 50% to 100% with a 10% interval were tested.

The opinion classification task is divided into two subtask: concern and sentiment classification. Concern classification is a multilabel text classification task, which classifies the comment sentences into one or more concern categories. Sentiment classification is a multiclass text classification task, which classifies the comment sentences into one concern category out of the three possible categories (supportive, neutral, and unsupportive). The convolutional neural network (CNN) algorithm was utilized for both concern and sentiment classification, because (1) it is the state-of-the-art algorithm for both multi-label and multi-class text classification, and (2) it does not require complicated feature engineering.

8.2.4 Implementation: Unsupervised Stakeholder Opinion Mining Method

8.2.4.1 Data Preparation

The stakeholder comment collection includes stakeholder comments on nine large-scale highway projects from eight states to ensure the coverage of different stakeholder concerns and opinions (see Table 7.2). The comment collection contains 3,132 comments, which were received during the projects' respective stakeholder involvement process, including comments provided through project websites, public hearings, emails, and social media. A total of 14,000 and 1,400 comment sentences out of the collection were randomly selected for the training and testing datasets.

To prepare the gold standard, the comment sentences in the testing dataset were manually analyzed and annotated. The gold standard annotations were determined based on mutual agreement among three annotators – the author, in addition to two researchers with expertise in stakeholder analysis and text classification. Table 8.2 shows examples of three comment sentences and their corresponding concern and sentiment category annotations.

Table 8.2 – Example Comment Sentences and their Concern and Sentiment Categories

Comment sentence	Concern categories	Sentiment category
This would create a greater alternative for people to travel our county, reduce air pollution, create long term jobs and help boost the economy without any business or residential acquisitions.	Mobility, air quality, regional development	Supportive
A review of the DEIS reveals that many of the stream crossings will be bridged, but only a select few are targeted for wildlife crossings.	Wildlife and habitat, physical infrastructure	Neutral
This project is completely unnecessary, and would do nothing to relieve the congestion and prevent accident.	Traffic, transportation safety	Unsupportive

8.2.4.2 Data Preprocessing

To implement the concern and sentiment labeling, the comment sentences were represented using the bag of words (BOW) model. Four commonly used preprocessing techniques were thus utilized: sentence splitting, tokenization, stopword removal, and lemmatization. Sentence splitting breaks a comment into sentences based on sentence boundary tokens, such as punctuations like ".","!", and "?". Tokenization then divides every comment sentence into meaningful units called tokens (e.g., words and punctuations), removes punctuations, and converts all words into their lowercase forms. Stopwords are those words (e.g., "to" "in", "on", and "the") that have high frequency but low discriminating power, which have little value in determining the category a comment sentence belongs to. By eliminating the nondiscriminative high-frequency words, stopword removal can reduce the number of features and reveal the discriminative words. However, stopword removal was only conducted for concern labeling, not for sentiment labeling. This is because certain stopwords can be good indicators of the commenter's sentiment, and removing them could completely change the sentiment of the comment. For example, in the comment sentence "I am not in favor of tolling existing roads", stopword "not" follows the opinion phrase "in favor of" and indicates that the commenter has unsupportive attitude. Lemmatization removes the inflectional endings of a word and returns its base or dictionary form, which is known as the lemma. By combining words with the same lemma, lemmatization can reduce the number of features, and can be effective in enhancing the classification performance. For example, after lemmatization, the words “supports”, “supported”, and “supporting” would all be transformed into their lemma “support”.

8.2.4.3 Algorithm Training and Testing

For concern and sentiment labeling, parameter tuning was conducted to optimize the labeling performance of a subset of training data. For example, the adapted LDA model has two important parameters: α, which controls the topic-document distribution, and β, which controls the topic-word distribution. To find the optimal α and β values, a range of values for each parameter (e.g., 0.01 to 1 with an interval of 0.01 for both α and β) was evaluated and the best combination of values (α = 0.07 and β = 0.67) was selected based on labeling accuracy.

For opinion classification, the CNN architecture by Kim (2014) was utilized, and the same word embeddings (that were used for concerns labeling) were used as the feature vectors. When developing the word embeddings, five was used as the size of the context window, and 300 was the dimension of the vector. The hyper-parameter tuning for the CNN classifiers were conducted using a combination of random and grid search based on the average classification performance (F1 measure) of 10-fold cross validation. For example, CNN has two important parameters: *alpha*, which is the learning rate that governs the weights update of each backpropagation; and *batch_size*, which controls the number of training data processed per gradient update. To select the optimal *alpha* and *batch_size,* a set of random combinations of the two parameters were evaluated first (e.g., select *alph* from 0.1 to 1 with an interval of 0.1 and select *batch_size* from 16 to 128 with an interval of 16). When the best combination (*alpha* = 0.9 and *batch_size* = 16) was identified, a fine-grained range of values (e.g,. 0.8 to 1.0 with an interval of 0.01 for *alpha*, and 16 to 32 with an interval of 1 for *batch_size*) were further tested. The optimal combination of values (*alpha* = 0.94 and *batch_size* = 32) was selected based on F1 measure. The LDA was implemented using the python topic modeling package genism (Rehurek and Sojka 2010), and the CNN classifier was implemented using the Python deep learning library Keras (Chollet 2015) and the Tensorflow (Abadi et al. 2016).

8.2.5 Evaluation

The performance of concern classification was evaluated using example-based multilabel evaluation metrics. Example-based precision and recall are calculated using Eqs. (8.10) and (8.11), where TP_i is the number of labels assigned correctly as positive for comment sentence d_j; FP_i is the number of labels assigned incorrectly for d_j; FN_i is the number of labels assigned incorrectly as negative for d_j; and *N* is the total number of testing sentences.

$$Example - based\ Precision = \frac{1}{N}\sum_{j=1}^{N}\frac{TP_j}{TP_j+FP_j} \tag{8.10}$$

$$Example - based\ Recall = \frac{1}{N}\sum_{j=1}^{N}\frac{TP_j}{TP_j+FN_j} \tag{8.11}$$

The performance of sentiment classification was evaluated using precision, recall, and F1 measure, as per Eqs. (8.12), (8.13), and (8.14), where true *TP* refers to the number of sentences classified correctly, *FP* refers to the number of sentences classified incorrectly, and *FN* refers to the number of sentences incorrectly classified as negative. Precision, here, is defined as the ratio of the number of correctly classified sentences over the total number of classified sentences. Recall, here, is defined as the ratio of the number of correctly classified sentences over the total number of sentences that should be classified. F1 is the harmonic mean of precision and recall. The aforementioned measures were calculated based on a comparison of the experimental results with the gold standard.

$$Precision = \frac{TP}{TP+FP} \tag{8.12}$$

$$Recall = \frac{TP}{TP+FN} \tag{8.13}$$

$$F1\ measure = \frac{2*Precision*Recall}{Precision+Recall} \tag{8.14}$$

8.3 Experimental Results and Analysis

A number of experiments were conducted to evaluate the effects of varying the size of the pseudo training data on the opinion classification performance. The proposed unsupervised method was then compared with existing supervised methods, and with the proposed tuple-based method (Chapters 6 and 7) in terms of classification performance.

8.3.1 Effect of Varying the Size of Pseudo Training Data

The effect of varying the size of the pseudo training data was evaluated. The results are summarized in Table 8.3. The best concern classification performance – 90.7%, 90.9%, and 90.8% exampled-based precision, recall, and F1 measure, respectively – was achieved using the top 90% of the pseudo training data. From 50% to 90%, the performance gradually improved despite the increase of noisy data (comment sentences mislabeled) in the training set. When using 100% of the pseudo training data, the performance did not decline significantly compared with the optimal point (90.5% vs 90.8%) in terms of F1 measure. The noisy data did not affect the performance too much because of the following two reasons: (1) the size of the noisy data is much smaller compared with the effective data in the training set; and (2) adding noisy data can prevent the CNN classifier from overfitting, thus minimizing the negative impact on the performance.

The best sentiment classification performance – 89.7%, 90.2%, and 89.9% precision, recall, and F1 measure, respectively – was achieved using the top 70% of the labeled comment sentences for training. Compared with concern classification, the optimal point for sentiment classification is lower (70% vs 90%), and the range of the F1 measure is smaller (88.3% - 89.9% vs. 86.8% - 90.8%), which indicates the sentiment labeling creates more noisy data compared with the concern labeling. Compared with concern classification, the standard deviation of the F1 measure across each category is smaller at every percentage of the pseudo training data used, which indicates that the sentiment classification has less variability in the performance across different categories.

Based on the aforediscussed experimental results, the proposed unsupervised method (1) uses LDA-based concern labelling and lexicon-based sentiment labelling for creating pseudo training data, and (2) trains the CNN algorithm on the top p% – 90% for concern classification and 70% for sentiment classification – of the labelled comment sentences.

Table 8.3 – Impact of Varying the Size of Pseudo Training Data on Classification Performance

Percentage	Precision	Recall	F1 measure	Standard deviation of F1 measure
Concern classification				
50%	86.4%	87.2%	86.8%	4.0%
60%	88.7%	89.2%	88.9%	3.7%
70%	89.5%	90.2%	89.8%	3.8%
80%	89.7%	90.4%	90.0%	4.2%
90%	90.7%	90.9%	90.8%	3.9%
100%	90.1%	90.8%	90.5%	4.1%
Sentiment classification				
50%	87.8%	88.8%	88.3%	1.3%
60%	88.5%	89.7%	89.1%	1.1%
70%	89.7%	90.2%	89.9%	0.9%
80%	89.4%	88.5%	88.9%	1.4%
90%	89.4%	88.4%	88.9%	1.2%
100%	88.2%	88.4%	88.3%	1.2%

8.3.2 Comparison with a Supervised Approach

A supervised approach was tested for comparison purposes. A support vector machines (SVM) model was trained on manually-annotated data, and BOW features, n-gram features, and semantic features developed from stakeholder concern lexicon and sentiment lexicon were utilized. SVM was selected for comparison because of its good performance indicated in previous opinion mining research (Sharma and Dey 2012; Zainuddin and Selamat 2014; Ravi and Ravi 2015). The performance results are summarized in Table 8.4. The proposed unsupervised method achieved a comparable level of performance (90.8% vs 92.0% and 89.9% vs 91.0% for F1 measure) for both concern and sentiment classification.

Table 8.4 – Comparison of the Proposed Opinion Mining Method with a Supervised Approach

Opinion mining method	Precision	Recall	F1 measure
Concern classification			
Proposed unsupervised method	90.7%	90.9%	90.8%
Existing supervised method	90.6%	93.5%	92.0%
Sentiment classification			
Proposed unsupervised method	89.7%	90.2%	89.9%
Existing supervised method	90.5%	91.5%	91.0%

8.3.3 Comparison with the Tuple-based Method

The proposed sentence-level method was then compared with the aforementioned tuple-based stakeholder opinion mining method (Chapters 6 and 7). To convert the tuple-based results to equivalent sentence-level results, for the sake of conducting an apple-apple comparison, the labels of all tuples in a comment sentence were aggregated to form the label set of the sentence. The performance of the two methods are summarized in Table 8.5. The results show that the proposed method achieved a better performance, on all metrics.

The Wilcoxon signed-rank test was further used to examine whether the better performance in F1 measure is significant across the 10-fold cross validation results. The Wilcoxon signed-rank test is a nonparametric test for comparing the differences between two-paired samples (Rey and Neuhäuser 2011). The result of the Wilcoxon signed-rank test was interpreted according to the probability value (p-value). The p-value is 0.0012, which is less than the 0.05 significance level. This indicates that there is a significant improvement in F1 measure when using the proposed sentence-level method. This indicates that, among the two methods, the sentence-level method is more suitable to use, if a sentence-level analysis is sufficient. If a more detailed, aspect-level analysis is desired, then the aspect-level method (Chapters 6 and 7) should be used.

Table 8.5 – Comparison of the Sentence-level Opinion Classification Method with the Tuple-based Method

Opinion mining method	Precision	Recall	F1 measure
Concern classification			
Proposed sentence-level method	90.7%	90.9%	90.8%
Proposed tuple-based method	86.7%	90.5%	88.6%
Sentiment classification			
Proposed sentence-level method	89.7%	90.2%	89.9%
Proposed tuple-based method	87.6%	87.3%	87.5%

8.3.4 Error Analysis

Two main types of errors were identified based on the testing results. First, comment sentences with implicit concerns were misclassified. For example, the following comment sentence should be classified into the “land use and property” category, because it expresses concerns over people whose properties would be displaced due to the project land use: “the thing that I hope that IDOT will consider is certainly treat all individuals who will be displaced in an extremely fair way”. However, because the sentence does not explicitly mention any words related to this category (e.g., “land”, “home”, or “property”), it was mistakenly classified into the “general concern” category. To address this type of error, some strategies could be considered and tested in future work. For example, more comment sentences with implicit concerns could be included in the training data and/or the concern seed words could be expanded to include more implicit terms/phrases. Second, comment sentences that express sentiment in an indirect or implicit way were misclassified. For example, the following sentence expresses supportive sentiment towards the project through double negatives, “opposing” and “short-sighted”, but was mistakenly classified into the unsupportive category: “anyone opposing these measures are short-sighted”. Similarly, the following sentence expresses unsupportive sentiment towards the toll road through a rhetorical question, but was mistakenly classified into the neutral category”: “how do you expect to get people to use the toll road when they could use Peotone Road for free?” In future work, syntactic and semantic features that represent such indirect or implicit sentiment expressions could be identified and integrated into the learning process, in order to test their effectiveness in dealing with such type of error.

CHAPTER 9: CASE STUDIES OF STAKEHOLDER OPINION MINING

9.1 Case Study Project Selection

The proposed stakeholder opinion mining method was used in analyzing stakeholder comments from three large-scale highway projects. The three projects are: the Illiana corridor project (Illinois and Indiana), the US 181 harbor bridge project (Texas), and the Chicago I-290 improvement project (Illinois). These particular projects were selected, because they all have significant impact on the surrounding environment, and their stakeholder comments are available to the public. For each project, two primary stakeholder groups were identified: agency and government, and individual and public organization. The agency and government stakeholder group includes planning agencies (e.g., FHWA, MPOs), resource agencies (e.g., Environmental Protection Agency, Department of Natural Resources), and the local government [e.g., the Cook County Board of Commissioners, the governing board and legislative body of the county, which consists of commissioners (elected officials)]. The individual and public organization stakeholder group includes any individual (e.g., a resident) or public organization (e.g., the Environmental Law and Policy Center, a Midwest-based non-profit environmental advocacy group) that has an interest in the proposed project, because they are affected by or have a concern about the project.

The Illiana corridor project was proposed as a toll way connecting northeast Illinois with northwestern Indiana. The planning process for Illiana corridor was conducted using a two-tier study. Tier-one study focused on identifying the transportation needs, developing and evaluating alternatives for all modes, and selecting a preferred corridor at the concept level. Tier-two study built on the selected preferred corridor, and conducted engineering analysis and environmental impact evaluation to identify specific design alternatives. Stakeholder comments were solicited during both studies. Because some comments from the tier-one study were used for training and testing, only comments received during the tier-two study were considered for the application study. The US 181 harbor bridge project includes the replacement of the existing harbor bridge (built in 1950s), and the reconstruction of portions of the US 181, the I-37, and the Crosstown Expressway. The construction of the new bridge began in 2016, and the whole project is expected to be completed in 2021. The Chicago I-290 improvement project was proposed to provide an improved transportation facility along the Eisenhower Expressway, which was initially constructed in the 1950s, and is now severely congested and accident-prone. The section of the Eisenhower Expressway that requires improvement is the primary corridor serving the travelers and the commuters in the greater Chicago area. Table 9.1 shows the number of comments received from the two stakeholder groups for the abovementioned three projects during their respective planning processes.

Table 9.1 – Statistics on Comments from Each Stakeholder Group of the Three Selected Projects

Project name	A & G*		I & P*		All stakeholders	
	# of comments	# of sentences	# of comments	# of sentences	# of comments	# of sentences
Illiana corridor tier 2	331	1,239	908	5,449	1239	6,688
US 181 Harbor bridge	52	246	103	741	155	987
Chicago I-290 improvement	36	272	271	1,170	307	1,442

* A & G=Agency and government; I & P = Individual and public organization

9.2 Stakeholder Opinion Mining Implementation

For the three projects, the proposed sentence-level opinion mining method (Chapter 8) was used to classify each comment sentence into one or more concern categories and into one support category. The concern and sentiment categories of each comment were determined based on the categories of all the sentences in that comment. The concern categories of a comment would be an aggregate of all the concern categories for each of its sentences. The sentiment category of a comment was determined based on the following heuristics: (1) if the comment contains one or more sentences with supportive opinions and other sentences with neutral opinions, and no sentences with unsupportive opinions, then the comment is categorized as supportive; (2) if the comment contains one or more sentences with unsupportive opinions and other sentences with neutral opinions, and no sentences with supportive opinions, then the comment is categorized as unsupportive; and (3) if the comment contains one or more sentences with supportive opinions and other sentences with unsupportive opinions, then the sentiment category of the whole comment is decided based on the majority vote.

For each project, the concern and sentiment classification results were analyzed to answer the following research questions, for each project: (1) what are the support levels of the stakeholders to the project? (2) What are the concerns of the stakeholders (i.e., the things that positively or negatively affect the stakeholders or are of interest or importance to them)? (3) What are the negative concerns of the stakeholders (i.e., the things that negatively affect the stakeholders or cause them worry or disturbance)? (4) What are the similarities and differences – in support levels, concerns, and negative concerns – across the different stakeholder groups?

9.3 Case Study Results and Analysis

9.3.1 The Support Levels of the Stakeholders

The distributions of the sentiments expressed by the stakeholder groups for the three projects are depicted in Figure 9.1. For the agency and government group, the majority of the comments are neutral (76%, 88%, and 89% for each project, respectively). For the individual and public organization group, the percentages of neutral comments are much lower (36%, 49%, and 58%). The difference in the sentiments of the two groups is likely due to their different roles and responsibilities in the transportation planning process. During the planning process, agency and government stakeholders are often responsible to collaborate with the lead agency (e.g., a state department of transportation) in tasks such as data and information collection and environmental impact analysis. They are thus more likely to provide recommendations to the project in the form of neutral comments. Individual and public organization stakeholders, on the other hand, have no obligation to corporate with the lead agency, and tend to provide comments when they have a strong opinion on the project. It is thus more likely to see more non-neutral (supportive or unsupportive) opinions in their comments.

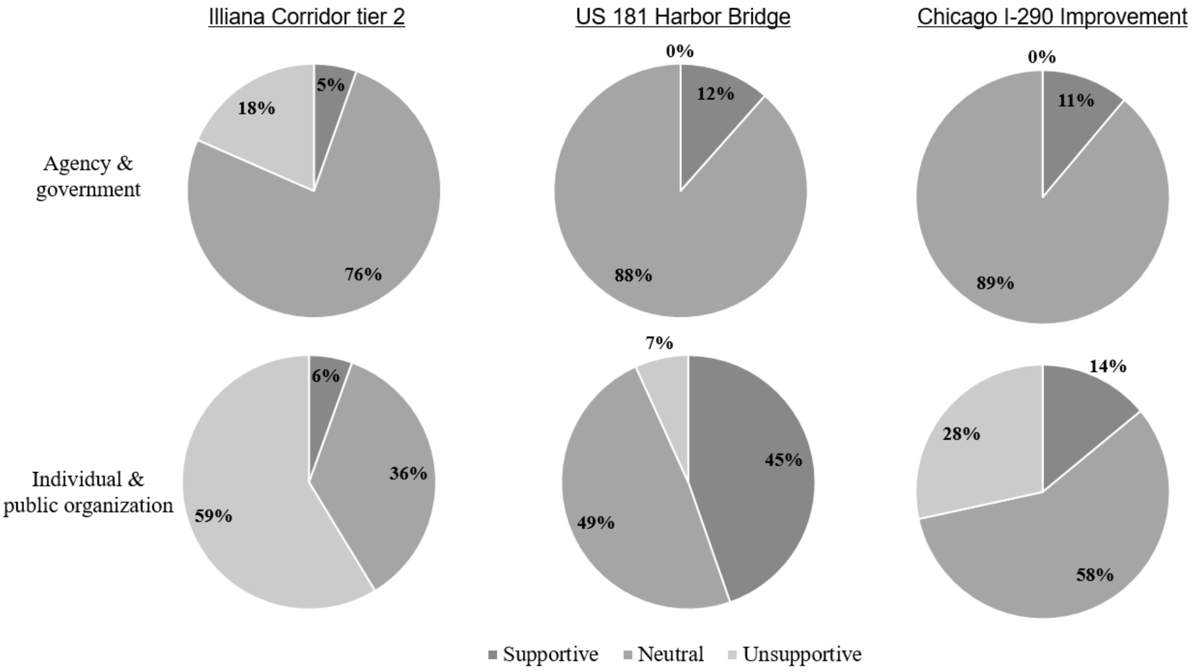


Figure 9.1 – Distributions of the Sentiments Expressed by the Stakeholder Groups for the Three Projects

The distribution of sentiments after removing the neutral comments is shown in Figure 9.2. For the Illiana corridor, most of the non-neutral comments are unsupportive for both stakeholder groups (77% and 91%), which indicates that the project received a consistently low level of stakeholder support during the planning process. This could mean that both groups have major negative concerns about the need, design concepts, or impacts of the project. For the US 181 Harbor Bridge, most of the non-neutral comments are supportive for both groups (100% and 87%), which indicates that the project received a consistently high level of stakeholder support during the planning process. This could mean that both groups agree on the need for the project, and on the proposed design concepts and mitigation measures. For the Chicago I-290 improvement project, all non-neutral comments from the agency and government group are supportive, but only 33% of the comments from the individual and public organization group are supportive. This indicates that the project did not receive a consistent level of support across the two groups. This could be due to the complex nature of the project, poor stakeholder communication issues, or simply differences in the views across both groups. These results indicate that in-depth planning studies and more stakeholder involvement activities may be needed for this project to allow both stakeholder groups to reach consensus or at least narrow differences in opinions.

The analysis also indicates that the opinions of the stakeholders, reflected in their comments, could be good predictors of the ultimate success or failure of a project. For the Illiana corridor, which could be safely categorized as a failing project [it was suspended in June 2015, with environmental issues causing the court to rule that the FHWA "erred in approving the project, because the project's environmental impact statement was the result of a "faulty" analysis" (Lafferty 2016)], 90% of the comments are unsupportive. For the US 181 Harbor Bridge, which so far seems to be successful (it passed the environmental review process and is currently under construction), 88% of the comments are supportive.

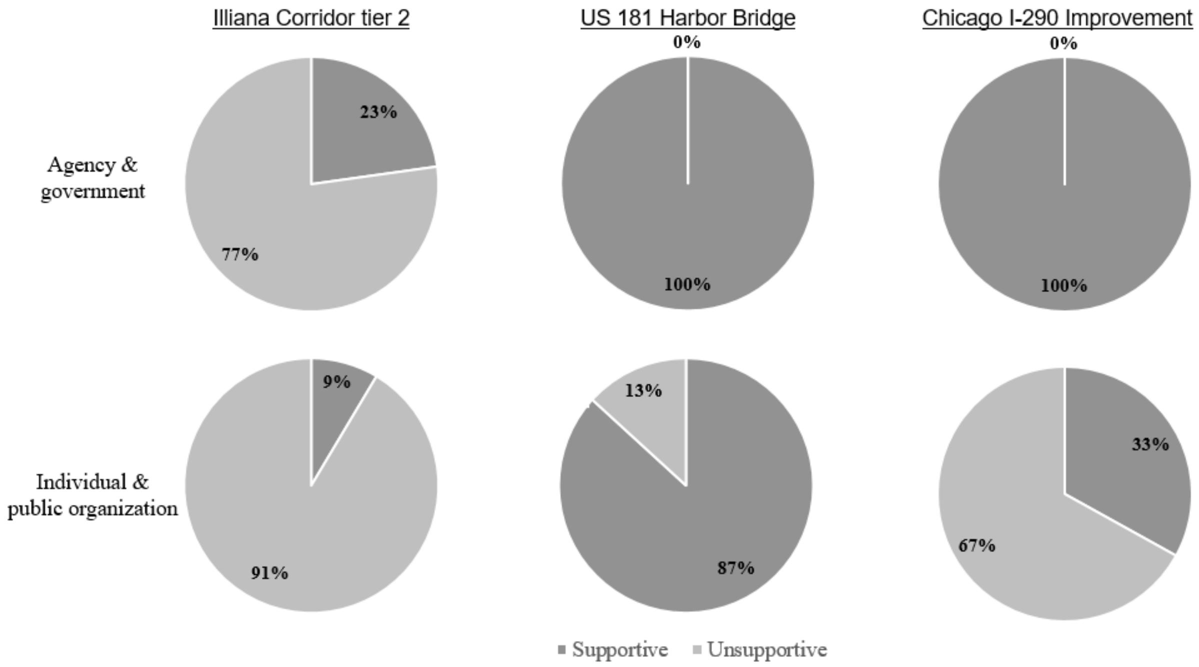


Figure 9.2 – Distributions of the Sentiments after Removing the Neutral Comments for the Three Projects

9.3.2 The Concerns of the Stakeholders

The distribution of the concerns by the stakeholder groups for the three projects are shown in Table 9.2. The top four most frequent concerns for each group, for each project, are highlighted in the table. As shown, for both groups, the majority of the concerns in the top-four list (i.e., 20 out of 24 highlighted values) are transportation and socioeconomic concerns. The impacts on the physical infrastructures is the top concern for both groups for both the Illiana corridor and the Chicago I-290 improvement project, and for the individual and public organization group for the US 181 Harbor Bridge. Other than that, almost all concerns appear in the top-four list – with the remaining ones coming close (in terms of the percentage of comments), which indicates that the stakeholders are, collectively, concerned about all environmental, transportation, socioeconomic, cultural, and management issues. For the agency and government group, the concerns in the top-four list are environmental, transportation, and socioeconomic concerns – in an equally distributed way (i.e., four in each category). For the individual and public organization group, the concerns in the top-four list are only transportation and socioeconomic concerns.

On average, for the three projects combined, the impacts on physical infrastructures and regional socioeconomic development are the most frequent concerns for both groups. Compared with the agency and government group, the individual and public organization group had more socioeconomic concerns (50% vs 31%), less environmental concerns (36% vs 68%), and comparable level of transportation concerns (51% vs 54%) in their comments. This is could be due to the different areas of expertise and interests of the two stakeholder groups. For example, the agency and government group includes several resource agencies (e.g., the Environmental Protection Agency), which have more expertise and interests in environmental issues than socioeconomic issues. The individual and public organization group often includes a larger number of stakeholders with more diverse backgrounds. Compared with the agency and government group, only a smaller percentage of stakeholders in this group [e.g., environmental nongovernmental organizations (NGOs)] have expertise and interests in environmental issues.

Table 9.2 – Distribution of Stakeholder Issues from All Stakeholder Comments on the Three Selected Projects

Concern category	Illiana corridor tier 2		US 181 Harbor Bridge		Chicago I-290 improvement		Average	
	A & G*	I & P*	A & G*	I & P*	A & G*	I & P*	A & G*	I & P*
Environmental concern								
Air quality	8%	12%	27%	14%	33%	10%	13%	12%
Water resource	43%	21%	10%	21%	28%	13%	37%	19%
Wildlife and habitat	40%	11%	15%	6%	17%	5%	35%	9%
Noise control	7%	5%	35%	11%	19%	9%	11%	7%
Subtotal	69%	38%	65%	42%	67%	30%	68%	36%
Transportation concern								
Traffic	14%	21%	33%	30%	17%	31%	17%	24%
Mobility and accessibility	8%	11%	29%	34%	25%	32%	12%	18%
Physical infrastructure	**44%**	**29%**	33%	**58%**	**53%**	**42%**	**43%**	**34%**
Transportation safety	4%	6%	10%	21%	3%	13%	5%	9%
Subtotal	54%	45%	54%	68%	61%	64%	54%	51%
Socioeconomic concern								
Cost and funding	6%	23%	13%	25%	17%	28%	8%	24%
Land use and property	10%	22%	38%	27%	19%	10%	14%	20%
Regional development	12%	23%	**52%**	43%	31%	19%	19%	23%
Subtotal	23%	52%	63%	50%	56%	41%	31%	50%
Cultural concern	9%	5%	17%	17%	19%	4%	11%	6%
Management concern	6%	6%	25%	6%	6%	7%	8%	7%
General concern	7%	11%	12%	17%	3%	10%	7%	11%

* A & G=Agency and government; I & P = Individual and public organization
For each stakeholder group, the top four most frequent concerns are highlighted in red color, and the most frequent concern is highlighted in bold

9.3.3 The Negative Concerns of the Stakeholders

The distribution of the negative concerns – concerns with unsupportive sentiment – of the stakeholder groups for the three projects are shown in Table 9.3. The top four most frequent negative concerns for each project are highlighted in the table. As shown, the majority of the negative concerns in the top-four list (i.e., 15 out of 16 highlighted values) are transportation and socioeconomic concerns, with an environmental concern only appearing in the list for the Illiana corridor project, for the agency and government group. Among these, the impacts on the physical infrastructures seems to be a common negative concern for both groups, for the three projects. The rankings of the subtypes of the negative concerns are, however, different from one project to another. For example the top negative concern for the agency and government group for the Illiana project is impacts on regional socioeconomic development, for the individual and public organization group is cost and funding, for the US 181 Harbor Bridge is a tie of four (impacts on regional socioeconomic development, land use and properties, physical infrastructures, and mobility and accessibility), and for the Chicago I-290 improvement project is impacts on physical infrastructures.

For the Illiana corridor, both groups shared similar negative concerns on the impacts on physical infrastructures, regional socioeconomic development, and land use and properties, with the individual and public organization group showing more concern for cost and funding and the agency and government group showing more concern for impacts on water resources (e.g., impact of the construction on the water quality).

On average, for the three projects combined, impacts on physical infrastructures, regional socioeconomic development, and land use and properties are the most frequent negative concerns for both groups. The results also indicate that, on average, stakeholders have little negative concerns about the impacts of the projects on cultural issues and about the process for managing the project.

Table 9.3 – Distribution of Stakeholder Concerns from Subjective Stakeholder Comments on the Three Selected Projects

Concern category	Illiana corridor tier 2		US 181 Harbor Bridge	Chicago I-290 improvement	Average	
	A & G*	I & P*	I & P*	I & P*	A & G*	I & P*
Environmental concern						
Air quality	7%	14%	0%	9%	7%	13%
Water resource	31%	22%	14%	12%	31%	20%
Wildlife and habitat	7%	13%	14%	3%	7%	12%
Noise control	3%	7%	0%	9%	3%	7%
Subtotal	39%	38%	29%	26%	39%	36%
Transportation concern						
Traffic	20%	20%	14%	38%	20%	23%
Mobility and accessibility	20%	8%	29%	36%	20%	12%
Physical infrastructure	28%	24%	29%	**40%**	28%	26%
Transportation safety	5%	5%	14%	14%	5%	6%
Subtotal	51%	41%	57%	65%	51%	45%
Socioeconomic concern						
Cost and funding	18%	**34%**	14%	32%	18%	**34%**
Land use and property	25%	26%	29%	14%	25%	25%
Regional development	**34%**	26%	29%	16%	**34%**	25%
Subtotal	48%	62%	57%	45%	48%	60%
Cultural concern	5%	5%	0%	4%	5%	4%
Management concern	3%	7%	0%	3%	3%	6%
General concern	7%	9%	0%	8%	7%	9%

* A & G=Agency and government; I & P = Individual and public organization

For each stakeholder group, the top four most frequent concerns are highlighted in red color, and the most frequent concern is highlighted in bold

CHAPTER 10: CONCLUSIONS, CONTRIBUTIONS, LIMITATIONS, AND RECOMMENDATIONS FOR FUTURE RESEARCH

10.1 Conclusions

10.1.1 Conclusions for Discovery of Integration Practices

A set of integration practices for integrating NEPA into statewide and metropolitan transportation project planning processes was discovered. The discovery of practices was based on literature review and two-tier expert opinion capturing. A set of potential integration practices were first identified based on a comprehensive literature review of existing integration guidelines and efforts and using expert input. The final set of practices were selected based on a survey of experts from relevant federal, state, and metropolitan planning, regulatory, and resource agencies. The selected practices were then integrated into existing transportation project planning processes and formalized into an integrated process flow with detailed implementation guidance. The integrated process was validated through a second expert survey.

As a result of the discovery efforts, two types of integration practices were identified and represented in the form of an integrated process flow with a detailed implementation guidance: process-oriented and collaboration-oriented integration practices. Process-oriented practices aim to allow for early and continuous agency participation; early identification of environmental, socioeconomic, and cultural impacts and concerns; reduced duplication of work; and reduced durations and efforts of project delivery. Collaboration-oriented practices aim to support the process-oriented practices by facilitating early, continuous, and in-depth interagency coordination and communication. The validation results indicate that the integrated process, including its implementation detail and process representation, provides appropriate guidance for integrating NEPA into transportation project planning processes in Illinois.

10.1.2 Conclusions for the Proposed Semantic Annotation Method and Algorithm

A domain-specific, deep semantic method for annotating documents in the transportation project environmental review (TPER) domain with functional process context concepts, which describe the subprocesses of the TPER process, was developed for supporting context-aware information retrieval in the TPER domain. The semantic analysis is facilitated through the use of a TPER epistemology, which is a semantic model that represents process, project, and resource contexts for supporting information retrieval. In developing the semantic annotation (SA) algorithm, a number of shallow and deep SA algorithms were proposed and tested on a testing data set of 1,328 Web pages in terms of mean precision (MP) and mean average precision (MAP) at the top 10, 20, 30, 40, and 50 documents.

For the shallow SA algorithms, the effects of syntactic concept expansion and filtering were investigated. It was found that syntactic concept expansion improves the overall performance of SA but brings a lot of noise at the same time. The results also showed that conducting syntactic concept filtering and domain-specific concept expansion after concept expansion is effective in enhancing performance through reducing the noise and through expanding the concept index with domain-specific concepts. The best-performing shallow SA algorithm, thus, includes syntactic concept expansion, syntactic concept filtering and domain-specific concept expansion, and syntactic term matching. It achieved 90% and 82% MP and MAP at the top 10 documents, respectively.

For the deep SA algorithms, eight different SS measures were tested. After comparing the performance of the shallow and deep methods, the best performance was achieved through a deep SA algorithm that uses the Al-Mubaid and Nguyen (2006) SS measure. The final SA algorithm, thus, includes semantic concept expansion, semantic similarity assessment using Al-Mubaid and Nguyen (2006) measure, and semantic term matching. This algorithm achieved 97% and 96% for MP and MAP at the top 10 documents, respectively. Compared with other SA work in other domains (Egozi et al. 2011; Fernandez et al. 2011), the proposed algorithm achieved high performance results (e.g., Egozi et al.'s and Fernadez et al.'s best MP values at the top 10 documents are 52.2% and 68%, respectively). The high performance can be mainly attributed to the use of: (1) a domain-specific concept space, which allowed for annotation based on specialized domain knowledge, (2) a contextualized concept representation, which facilitated annotation using contextual information, (3) semantic concept expansion for incorporating semantically related concepts in the annotation process, which provided a more complete concept space, and (4) semantic similarity measures for assessing the match between the original concept and expansion concepts, which led to more accurate annotation weights.

10.1.3 Conclusions for the Proposed Context-aware Information Retrieval Method and Algorithm

A new context-based relevance assessment method was developed, which allows for enhanced context representation through two proposed approaches: (1) a context-aware and deep semantic concept indexing approach, and (2) a deep and semantically-sensitive relevance estimation approach. Accordingly, two context-enhanced document ranking methods were proposed: (1) a context-enhanced vector space model (VSM)-based method, which uses context similarity to measure the relevance of a document to a query based on the similarity between their contextual concepts, and (2) a context-enhanced statistical language model (SLM)-based method, which uses context probability to measure the relevance of a document to a query based on the likelihood that the document is relevant to a query on the contextual level.

Based on a testing data set of 5,436 Web pages and 18 queries, the context-enhanced VSM-based method outperformed the context-enhanced SLM-based method on every performance metric. The context-enhanced VSM-based method achieved 48% MAP, and 79%, 70%, 68%, 66%, and 65% MP at the top 10, 20, 30, 40, and 50 retrieved documents, respectively. Compared to the keyword-based VSM method, the results also showed that the integration of the proposed context-based relevance assessment method is effective in improving information retrieval performance, and can effectively deal with the performance drop due to the increase of query length.

10.1.4 Conclusions for the Proposed Stakeholder Opinion Extraction Method and Algorithm

A domain-specific, supervised ML-based information extraction method for extracting subject, concern, and opinion expressions from stakeholder comments on large-scale transportation projects, was developed, for supporting stakeholder opinion mining. This method would facilitate the early identification of stakeholder concerns and support by eliminating manual efforts and enabling a broader stakeholder outreach. The method has the potential to significantly improve the efficiency of the stakeholder involvement process in terms of time and cost. In developing the proposed method, several supervised machine learning algorithms were tested and evaluated, and the effects of using dependency features and semantic features (including two domain-specific semantic features) were also studied. To further improve the recall of the information extraction results, a set of language rules based on linguistic patterns were developed and were combined with the selected machine learning algorithm and features. All the methods/algorithms were tested on a testing data set of 440 comment sentences, which were selected from a comment collection including 3,112 stakeholder comments on nine large-scale transportation projects. Based on the experimental results, the final proposed method uses a linear-chain CRF algorithm for learning; syntactic, dependency, and semantic features for characterizing the text; and language rules for supporting the extraction. The proposed method achieved 93% precision, 89% recall, and 91% F1 measure on the testing data.

10.1.5 Conclusions for the Proposed Stakeholder Opinion Classification Method and Algorithm

A domain-specific, unsupervised ML-based stakeholder opinion classification method for identifying the concerns and support levels of stakeholders during the early stage of highway project decision making was developed. The proposed method classifies the aspect-level opinion tuples from stakeholder comments into different concern categories (e.g., mobility and accessibility, air quality, transportation safety, etc.) and into one sentiment category (supportive, unsupportive, or neutral). The proposed method can automatically create labeled training data through iteratively generating clusters of opinion tuples, based on keywords, for each classification category. For clustering, semantic similarities between opinion tuples are captured through opinion semantic vectors, which are learned from a domain-specific text corpus using the skip-gram word-embedding model. An adapted k-means algorithm is then used for clustering. The top *p*% of opinion tuples in the clusters are then used for training a supervised ML algorithm in concern and sentiment classification. Overall, the proposed method achieved 88%, 90%, and 89% exampled-based precision, recall, and F1 measure, respectively, for concern classification; and 87%, 86%, and 86% precision, recall, and F1 measure, respectively, for sentiment classification. Compared to existing supervised methods, the proposed method achieved a comparable level of classification performance – but without any need for manual training data labeling.

10.1.6 Conclusions for the Proposed Sentence-level Stakeholder Opinion Mining Method and Algorithm

A domain-specific, unsupervised machine learning-based stakeholder opinion mining method to identify the concerns and support levels of the stakeholders, from their comments, during the highway planning process, was developed. Compared to the tuple-based method (Section 10.1.4 and 10.1.5), the sentence-level method offers an alternative approach when a sentence-level analysis is sufficient. The proposed sentence-level method can automatically create pseudo training data through LDA-based concern labeling and lexicon-based sentiment labeling. The convolutional neural network (CNN) algorithm is then trained on a subset of the automatically-labelled comment sentences, which is selected based on the concern and sentiment confidence scores of the sentences. Overall, the proposed method achieved 90.7%, 90.9%, and 90.8% example-based precision, recall, and F1 measure, respectively, for concern classification; and 89.7%, 90.2%, and 89.9% precision, recall, and F1 measure, respectively, for sentiment classification. Compared to the tuple-based method (Section 10.1.4 and 10.1.5), the proposed method achieved a higher level of classification performance – but it could do the analysis on the sentence-level only.

10.1.7 Conclusions for the Case Studies of Stakeholder Opinion Mining

The implementation of the proposed stakeholder opinion mining method to analyze stakeholder comments on three large-scale highway projects shows how different stakeholder groups could display different concerns and levels of support because of their different roles and areas of expertise or interest. For example, for the three projects combined, the individual and public organization group had more socioeconomic concerns but less environmental concerns compared with the agency and government group. It also shows how different stakeholder groups could share similar concerns and agree on supporting (or opposing) a project. For example, impacts on physical infrastructures was a common negative concern for both groups, for the three projects. The analysis also indicates that the opinions of the stakeholders, reflected in their comments, could be good predictors of the ultimate success or failure of a project.

10.2Contributions to the Body of Knowledge

10.2.1 Contributions for Discovery of Integration Practices

This research contributes to the body of knowledge in three main ways. First, it identifies a set of context-sensitive integration practices for environmental streamlining to improve project delivery without compromising environmental compliance; through an in-depth investigation of existing planning processes in Illinois and a thorough assessment of potential integration practices based on the opinion of federal, state, and local experts, the practices were adapted to the state of Illinois context. Second, this research models the set of integration practices in the form of an integrated process model and provides well-defined guidance on the implementation and evaluation of the integrated process; a well-defined process flow along with a textual description supports both clarity and detail in process description. The integrated process advances the knowledge in the area of environmental streamlining area by (1) incorporating NEPA with transportation planning at both the system level and the corridor level, (2) providing context-specific implementation detail on how to conduct environmental analysis during the planning process, and (3) establishing standardized/formalized performance measures to evaluate the implementation of the integrated process. Third, this research offers a methodology for process streamlining based on case study review and two-tier expert opinion capturing; future process streamlining efforts in the construction domain could benchmark this methodology. The implementation of the integrated process would improve interagency coordination and communication, enable early identification of potential environmental issues and early consideration of avoidance/mitigation measures, and facilitate the use of early planning data/decisions in subsequent NEPA studies; all would result in improving the decision-making process, reducing duplication of work, and enhancing project delivery in terms of time and cost.

10.2.2 Contributions for the Proposed Semantic Annotation Method and Algorithm

This research offers a domain-specific, deep semantic annotation (SA) method for annotating documents in the TPER domain with process context concepts using a TPER epistemology. This would facilitate context-aware information retrieval in the TPER domain, because it enables domain-specific, and thus more accurate, automated semantic annotations.

Beyond this application, this work additionally contributes to the body of knowledge in six main ways. First, this research offers a TPER epistemology for supporting context-aware information retrieval in the TPER domain. The TPER epistemology is a formal representation of the knowledge in the TPER domain, which can support context-aware information retrieval through semantic annotation, semantic query processing, and semantic document ranking. Second, this research offers a baseline domain-specific, deep SA method for annotating documents with concepts in the TPER epistemology. This algorithm could serve as a benchmark for future research and could provide opportunities for adaptation to annotate other types of documents with other concepts in the TPER epistemology or in other transportation domain semantic models. Third, this research shows the effectiveness of syntactic concept expansion and filtering for shallow SA. The experimental results indicate that concept expansion through WordNet can improve the overall SA performance but could also bring noise. The results further show that concept filtering and domain-specific concept expansion are effective in removing the noise and in expanding concept terms with context terms, both which result in enhanced performance. Fourth, this research provides a comparison of shallow and deep SA. The experimental results show that deep SA methods outperform shallow SA methods. Fifth, this research provides a comparison of different semantic similarity (SS) measures for deep SA. Eight SS measures were experimentally tested. The results show that Al-Mubaid and Nguyen (2006) SS measure achieved the best performance. Sixth, this work offers a dataset of annotated Web pages for the TPER domain. This dataset can serve as the gold standard for future researchers to evaluate SA algorithms for the TPER domain.

10.2.3 Contributions for the Proposed Context-aware Information Retrieval Method and Algorithm

This research offers a new context-based relevance assessment method to support context-enhanced document ranking for retrieving relevant documents in the TPER domain. The proposed context-enhanced document ranking method would improve the ability of transportation practitioners to find the right information, at the right time, for the task at hand; this would help support project decision making and would reduce the time that agency employees spend to look for information in unstructured documents.

Beyond this application, this work additionally contributes to the body of knowledge in three main ways. First, this research offers a new context-based semantic relevance assessment method to enrich both the domain-specific representation of context and the contextual information considered for enhanced document relevance recognition. The proposed method improves the existing state-of-the-art methods from the following two perspectives: (1) it provides an enhanced and deep representation of context by using a domain-specific context model and extending the original concept terms with concept terms from semantically-related concepts; and (2) it achieves deeper level and semantically-sensitive relevance assessment by representing the original query through a semantically-extended set of concepts and considering their relative semantic relatedness to differentiate their level of relevance to the original query.

Second, this research compares the vector space model (VSM) and the statistical language model (SLM) in context-enhanced semantic document ranking in the TPER domain. To enable context-based semantic document ranking, this research proposes the use of context similarity and context probability to integrate the proposed context-based semantic relevance assessment into the VSM and the SLM, respectively. The experimental results show the effectiveness of the integration from two perspectives: (1) using context similarity and context probability can significantly improve the overall information retrieval performance of keyword-based methods, and (2) using context similarity and context probability is effective in dealing with the performance drops due to the increase of query length. When comparing the two context-enhanced methods, the experimental results show that the context-enhanced VSM-based method outperforms the context-enhanced SLM-based method.

Third, this work offers a dataset of manually judged Web pages and queries for the TPER domain. This dataset can serve as an experimental corpus for future researchers to evaluate information retrieval approaches in the same domain.

10.2.4 Contributions for the Proposed Stakeholder Opinion Extraction Method and Algorithm

This research offers a baseline domain-specific, supervised ML-based information extraction method for extracting subject, concern, and opinion expressions from stakeholder comments on large-scale transportation projects. This method would facilitate the early identification of stakeholder concerns and support during the transportation project development process.

Beyond this application, this work additionally contributes to the body of knowledge in sixth main ways. First, this method could serve as a benchmark for future research and could provide opportunities for adaptation to extract other useful information from stakeholder comments on transportation projects or other types of infrastructure projects. Second, this research evaluates the performance of five supervised machine learning algorithms in information extraction, particularly in opinion extraction from stakeholder comments on large-scale infrastructure projects. The experimental results indicate that the linear-chain conditional random fields (CRF) algorithm achieves the best performance. Third, this research proposes and develops a stakeholder concern hierarchy and a key phrase list to better capture semantic features of the text. Fourth, this research evaluates the impact of dependency and semantic features (including two domain-specific semantic features) on the performance of information extraction. The experimental results show that both dependency and semantic features can enhance the performance of information extraction in terms of precision and recall. Fifth, this research offers a set of language rules to improve the recall of information extraction when combined with the linear-chain CRF and the syntactic, dependency, and sematic features. The experimental results show that despite a slight decrease in precision, the use of language rules could improve the recall of information extraction and the overall F1 measure. Sixth, this work offers a dataset of labeled stakeholder comment sentences that could serve as a gold standard for future researchers to evaluate information extraction methods.

10.2.5 Contributions for the Proposed Stakeholder Opinion Classification Method and Algorithm

This research offers a domain-specific, unsupervised ML-based stakeholder opinion classification method for identifying the concerns and support levels of stakeholders during the early stage of highway project decision making. The proposed method would help identify the sentiments and concerns of stakeholders through instant and automatic recognition of concerns and support levels from stakeholder opinions. The method would enable a broader public outreach through the consideration of comments from social media, and would enhance the decision makers' ability to proactively resolve issues before they escalate into bigger problems.

Beyond this application, this research additionally contributes to the body of knowledge in five primary ways. First, this research offers an unsupervised machine learning-based opinion classification method that can automatically create labeled training data based on only keywords for each classification category. The proposed method captures the semantic similarity between opinion tuples through representing the tuples as opinion semantic vectors. It also adapts the k-means clustering algorithm to incorporate semantic similarity and the characteristics of both concern clusters and sentiment clusters. Second, this research evaluates the performance of four different opinion semantic vectors in opinion classification. The experimental results indicate that the best performance, in both concern and sentiment classification, is achieved when learning the opinion semantic vectors from the domain-specific comment collection using the skip-gram word-embedding model. Third, this research compares the performance of the support vector machines (SVM) algorithm and the backpropagation for multilabel learning (BP-MLL) algorithm in multilabel concern classification, when trained on the automatically-generated training data. The experimental results indicate that the SVM algorithm achieves better performance. Fourth, this research investigates the impact of varying the percentage of clustered opinion tuples used for training. For the given type of text and categories, the results indicate that the optimal percentages for concern and sentiment classification are in the range of 80% and 70%, respectively. Fifth, this research compares the performance of the proposed unsupervised opinion classification method with existing supervised methods. The experimental results show that the proposed unsupervised method can achieve a comparable level of classification performance, while saving the manual effort in labeling.

10.2.6 Contributions for the Proposed Sentence-level Stakeholder Opinion Mining Method and Algorithm

This research offers a domain-specific, unsupervised machine learning-based stakeholder opinion mining method for classifying comment sentences on large-scale highway projects into one or more concern categories, and into one sentiment category. Transportation planners could use this method to automatically identify the concerns and support levels of the stakeholders from their comments, which could allow for a broader and more diverse public outreach and could improve the accessibility of the stakeholders to the transportation planning decision making process and the responsiveness of the process to the stakeholders' concerns.

Beyond this application, this research additionally contributes to the body of knowledge in four primary ways. First, the research proposes an unsupervised ML-based stakeholder opinion mining method that can automatically create pseudo training data through LDA-based concern labeling and lexicon-based sentiment labeling. For concern labelling, the proposed method adapts the LDA model and the collapse Gibbs sampling method through integrating the pre-defined seed words and semantic similarities into topic-assignment and topic-word distributions. For sentiment labeling, a lexicon-based method is proposed to aggregate word-level sentiments and word-negation relations to estimate the sentence-level sentiment confidence scores. Second, this research investigates the impact of varying the size of the pseudo training data on the classification performance. For the given type of text and categories, the results indicate that the optimal percentages for concern and sentiment classification are in the range of 90% and 70%, respectively.

Third, this research compares the performance of the proposed unsupervised method with the supervised approach. The experimental results show that the proposed method can achieve a comparable level of classification performance, while saving the manual effort in labeling. Fourth, this research compares the proposed sentence-level stakeholder opinion mining method with the proposed tuple-based method (Section 10.2.4 and 10.2.5). The experimental results show that the sentence-level method achieved higher performance, which indicates that, among the two methods, the sentence-level method is more suitable to use, if a sentence-level analysis is sufficient.

10.2.7 Contributions for the Case Studies of Stakeholder Opinion Mining

The use of the proposed stakeholder opinion mining method to analyze stakeholder comments on three large-scale highway projects provides a better understanding of stakeholder opinions, and how they could be similar or different across different stakeholder groups. The analysis also indicates that the opinions of the stakeholders, reflected in their comments, could be good predictors of the ultimate success or failure of a project.

10.3Limitations

Seven main limitations of the work are acknowledged. First, for discovering the integration practices, the number of transportation practitioners that were involved in the two-tier survey to develop and validate the integrated process was limited. In future work, another validation study could be conducted to involve more transportation practitioners in the validation of the integrated process.

Second, for the semantic annotation work, the proposed method focused on annotating documents with only functional process context concepts, and was evaluated on a small collection of documents. In future work, the proposed semantic annotation method could be extended to use all the concepts in the semantic model for annotation, and a larger document collection could be used for evaluation.

Third, for the context-aware information retrieval work, the evaluation was conducted using a limited document collection and a limited set of testing queries. Also, the queries were developed only using a qualitative method (expert interviews), and the experts (industry practitioners) who helped identify the testing queries were not involved in the relevance judgment. Unlike other general information retrieval research efforts, where standard document collections and queries are commonly used, this work focuses on the transportation environmental review domain, where standard collections and queries are not available. Although some variability in the performance may occur if the information retrieval methods are evaluated using different datasets and queries, a similar performance is expected if the text exhibits similar semantic features. In future research, an information retrieval testing system – that allows actual users to provide their own queries and select relevant documents based on their own information needs – could be developed and used for improved testing and evaluation.

Fourth, for the stakeholder opinion extraction work, because of being supervised, the proposed method requires manual labeling of a large amount of stakeholder comments for training. In future work, an unsupervised or semi-supervised method could be developed to reduce such manual effort.

Fifth, for the stakeholder opinion classification work, the proposed method used the same global threshold for all the categories when selecting the size of pseudo training data. In future research, more flexible ways of selecting the optimal size of pseudo training data could be explored, such as using a set of per-category thresholds.

Sixth, for the sentence-level stakeholder opinion mining work, the proposed method can only identify coarse-grained sentiments (supportive, neutral, and unsupportive) from stakeholder comments, and has limited capabilities in detecting non-typical, complicated opinions. In future work, the proposed method could be extended to conduct finer-grained sentiment analysis and handle comments with ambiguous or mixed opinions.

Seventh, for the opinion mining case studies, the studies used all stakeholder comments that were received during the projects' public comment periods, without filtering or analysis of influences or potential biases. For example, a single stakeholder could provide multiple comments, which could introduce bias or amplify certain concerns over the others. Also, there is no guarantee that the comments are representative of all stakeholder opinions, especially that most of the comments were either submitted during the public hearings or online – for example senior stakeholders may have difficulties in attending public hearings or commenting online. In future work, the proposed stakeholder opinion mining methods could be integrated with a stakeholder analysis to identify and filter potential biases in the comments, assess the representativeness of the comments, and further analyze the profiles of the stakeholders to identify the potential differences in the influences of their opinions.

10.4Recommendations for Future Research

Future research is recommended in four main directions. First, a GIS-based, natural language processing (NLP)-enabled, semantic system for environmental review and management could be developed and used to further support the streamlining of environmental and project development processes. Current research efforts towards the use of NLP and semantic analyses in GIS systems are becoming increasingly important (Lampoltshammer 2012). Three primary research paths can be followed: (1) using NLP-enabled interfaces, in addition to the traditional visual interfaces, to support enhanced interaction between users and the GIS system based on natural language, for example through queries; (2) using information retrieval techniques to complement GIS information with textual information from the World Wide Web to support environmental review, for example through retrieving relevant information about potential mitigation measures used for other projects and in other states; and (3) combining GIS with semantic reasoning to support enhanced environmental decision making, for example through suggestion-making functions.

Second, an information retrieval system could be developed to further validate the proposed context-aware information retrieval method and investigate its impact on transportation decision making. The system could be used to solicit actual search queries from transportation practitioners, and allow users to select relevant documents based on their own information needs. It could also allow further optimization of the proposed information retrieval method based on user feedback. The system could also be used to conduct further case studies to better understand the impact of the proposed information retrieval approach on the efficiency of domain-specific information-seeking tasks and the overall transportation decision making process.

Third, in future work, an infrastructure project decision making support system could be developed to facilitate real-time stakeholder involvement. Throughout the project planning and development process, stakeholders could use the system to find project-specific information, such as the project’s feasibility study and environmental study, submit comments, and express their sentiments towards the project in terms of rating (e.g., four out of five stars). The system could help identify concerns from stakeholder comments and prioritize comments for immediate response in a real-time manner. As the project progresses, the system could monitor the trends of stakeholder concerns and sentiments, evaluate the project’s responses to the stakeholder concerns, and provide recommendations for project planning and design.

Fourth, further research is recommended to integrate stakeholder opinion mining with social network analysis and stakeholder analysis to analyze the comments received at different timeframes of the project planning process. Time-series analyses could be conducted to evaluate the dynamics of the stakeholder opinions and investigate how opinions are affected by key events and by opinions from other stakeholder groups. Influential events and stakeholders could be identified to provide recommendations for facilitating effective and efficient stakeholder involvement. Models could be developed to predict future stakeholder concerns and levels of support given current stakeholder opinions and future scenarios.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., and Kudlur, M. (2016). "TensorFlow: A System for Large-Scale Machine Learning." *OSDI*, 16, 265-283.

Abbasi, M. K., and Frommholz, I. (2015). "Cluster-based polyrepresentation as science modelling approach for information retrieval." *Scientometrics*, 102(3), 2301-2322.

Aggarwal, C. C., and Reddy, C. K. (2013). *Data clustering: algorithms and applications*, CRC press, Boca Raton, FL.

Aggarwal, C. C., and Zhai, C. (2012). *Mining text data*, Springer Science and Business Media, Berlin, Germany.

Akhtar, M. S., Gupta, D., Ekbal, A., and Bhattacharyya, P. (2017). "Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis." *Knowl.-Based Syst.*, 125, 116-135.

Alavi, M., and Leidner, D.E. (2001). "Review: knowledge management and knowledge management systems: conceptual foundations and research issues." *MIS Quarterly*, 25(1), 107-136.

Alghunaim, A., Mohtarami, M., Cyphers, S., and Glass, J. (2015). "A Vector Space Approach for Aspect Based Sentiment Analysis." *Proc., of NAACL-HLT 2015, 116-122.*

Allan, J., Aslam, J., Belkin, N., Buckley, C., Callan, J., Croft, B., Dumais, S., Fuhr, N., Harman, D., Harper, D.J., Hiemstra, D., Hofmann, T., Hovy, E., Kraaij, W., Lafferty, J., Lavrenko, V., Lewis, D., Liddy, L.,Manmatha, R., McCallum, A., Ponte, J., Prager, J., Radev, D., Resnik, P., Robertson, S., Rosenfeld, R., Roukos, S., Sanderson, M., Schwartz, R., Singhal, A., Smeaton, A., Turtle, H., Voorhees, E., Weischedel, R.,Xu, J., and Zhai, C.X. (2003). “*Challenges in information retrieval and language modeling: report of a workshop held at the center for intelligent information retrieval, University of Massachusetts Amherst, September 2002.” ACM SIGIR Forum*, 37(1), 31-47.

AlMasri, M., Tan, K., Berrut, C., Chevallet, J. P., and Mulhem, P. (2014). “Integrating semantic term relations into information retrieval systems based on language models.” *Asia Information Retrieval Symposium*, 136-147.

Al-Mubaid, H., and Nguyen, H. A. (2006). “A combination-based semantic similarity measure using multiple information sources.” *IEEE Int. Conf. Inform. Reuse and Integr.*, IEEE, Piscataway, NJ, 617-621.

Alsubaey, M., Asadi, A., and Makatsoris, H. (2015). “A naive bayes approach for EWS detection by text mining of unstructured data: a construction project case.” *IEEE IntelliSys* 2015, IEEE, Piscataway, NJ, 164-168.

Altun, Y., Tsochantaridis, I., and Hofmann, T. (2003). “Hidden Markov support vector machines.” *Proc., 20th Intl. Conf. on Machine Learning*, 3-10.

Anjaria, M., and Guddeti, R. M. R. (2014). "Influence factor based opinion mining of Twitter data using supervised learning." *2014 Sixth Intl. Conf. on Communication Systems and Networks (COMSNETS),* IEEE, Piscataway, NJ, 1-8.

Appel, O., Chiclana, F., Carter, J., and Fujita, H. (2016). “A hybrid approach to the sentiment analysis problem at the sentence level.” *Knowl.-Based Syst.*, 108, 110-124.

Babashzadeh, A., Huang, J., and Daoud, M. (2013). “Exploiting semantics for improving clinical information retrieval.” *Proc., 36th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, ACM, New York, NY, 801-804.

Baccianella, S., Esuli, A., and Sebastiani, F. (2010). “SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.” *LREC*, 10, 2200-2204.

Bahrami, A., Yuan, J., Smart, P.R., and Shadbolt, N. R. (2007). “Context aware information retrieval for enhanced situation awareness.” *Military Comm. Conf. 2007*, IEEE, Piscataway, NJ, 1-6.

Barberio, G., Barolsky, R., Culp, M., and Ritter, R. (2008a). “PEL – a path to streamlining and stewardship.”a *Public roads*, <http://www.fhwa.dot.gov/publications/publicroads/08mar/01.cfm> (March 23, 2014).

Barberio, G., Barolsky, R., Culp, M., and Ritter, R. (2008b). “Planning and environmental linkages (PEL): using the PEL umbrella approach to streamline transportation decision making.” *J. Transp. Research Board*, 2058, 1-13.

Baroni, M., Dinu, G., and Kruszewski, G. (2014) “Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors.” *Proc., 52th Ann. Meeting of Assoc. for Comput. Linguist.* 238-247.

Basu, A., Walters, C., and Shepherd, M. (2003). “Support vector machines for text categorization.” *Proc. 36th Annual Hawaii International Conference on System Sciences*,

Bennett, G., Scholer, F., and Uitdenbogerd, A. (2008). “A comparative study of probabilistic and language models for information retrieval.” *Proc. 19th Conf. Australasian Database*, 75, 65-74.

Bikakis, N., Giannopoulos, G., Dalamagas, T., and Sellis, T. (2010). “Integrating keywords and semantics on document annotation and search.” *Proc. On the Move to Meaningful Internet (OTM) Systems*, 921-938.

Blei, D. M., Ng, A.Y., and Jordan, M.I. (2003). "Latent Dirichlet allocation." *J. Machine Learning Research*, 3(1), 993-1022

Boubekeur, F., Boughanem, M., Tamine, L., and Daoud, M. (2010). "Using WordNet for concept-based document indexing in information retrieval."*Proc., 4th Int. Conf. Semant. (SEMAPRO),* Int. Academy, Research, and Industry Assoc. (IARIA), NY, 151-157.

Bouramoul, A., Kholladi, M. K., and Doan, B. L. (2012). "An ontology-based approach for semantics ranking of the web search engines results." *Proc. 2012 Intl. Conf. on Multimedia Computing and Systems (ICMCS),* 797-802.

Budanitsky, A., and Hirst, G. (2006). "Evaluating wordnet-based measures of lexical semantic relatedness." *Comput. Linguist.*, 32(1), 13-47.

Buthelezi, M., and Mkhize, P. (2014). "Factors influencing quality of knowledge shared in software development community of practice." *Proc., 11th Int. Conf. Intellectual Capital, Knowl. Manage. and Org. Learning (ICICKM2014).* Academic Conference and Publishing International Limited (ACPIL), Reading, UK, 91-100.

Buttcher, S., Clarke, C. L., and Cormack, G. V. (2010). *Information retrieval: Implementing and evaluating search engines*, Mit Press, Cambridge, MA.

Cambridge Systematics, Inc. (2011). *Accelerating Federal Program and Project Delivery,* Orange County Transportation Authority, Orange, CA.

Cambridge Systematics, Inc. (2013). *NCHRP report 754: Improving management of transportation information,* TRB, Washington, D.C., 43-60.

Castells, P., Fernandez, M., and Vallet, D. (2007). "An adaptation of the vector-space model for ontology-based information retrieval." *IEEE Trans. Knowl. Data Eng.*, 19(2), 261-272.

Ceri, S., Bozzon, A., Brambilla, M., Della Valle, E., Fraternali, P., and Quarteroni, S. (2013). *Web Information Retrieval*, Springer Science and Business Media, Berlin, Germany.

Chauhan, R., Goudar, R., Sharma, R., and Chauhan, A. (2013). "Domain ontology based semantic search for efficient information retrieval through automatic query expansion." *Proc., 2013 Int. Conf. Intell. Syst. and Sig. Proc.*, IEEE, Piscataway, NJ, 309-402.

Chen, Z., Mukherjee A., and Liu B. (2014). "Aspect extraction with automated prior knowledge learning." *Proc., of ACL,* 347-358.

Chen, Q., Zhu, W., Ju, C., and Zhang, W. (2014). "Cross domain web information extraction with multi-level feature model." *Proc., of 10th ICNC,* 780-784.

Choi, Y., and Cardie, C. (2008). "Learning with compositional semantics as structural inference for subsentential sentiment analysis." *Proc. Conf. on Empirical Methods in NLP (EMNLP 08),* 793-801.

Chollet, F. (2015). "Keras", < **https://keras.io**>.

Choudhary, A. K., Oluikpe, P. I., Harding, J. A., and Carrillo, P. M. (2009). "The needs and benefits of Text Mining applications on Post-Project Reviews." *Computers in Industry*, 60(9), 728-740.

Clark, E. R., and Canter, L. W. (1997). *Environmental policy and NEPA: past, present, and future*, CRC Press, Boca Raton, FL, 47-62.

Council on Environmental Quality (CEQ). (2006). "NEPA's forty most asked questions." <http://ceq.hss.doe.gov/nepa/regs/40/1-10.HTM> (August 23, 2012).

Collins, M. (2002). "Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms." *Proc., ACL-02 Conf. on Empirical Methods in NLP*, 10, 1-8.

CEQ. (2007). “Implementing the NEPA process.” *A citizen’s guide to the NEPA,* CEQ, Washington, DC, 10-21.

De Jong, T. (1996). “Types and qualities of knowledge.” *Edu. Psychologist,* 31(2), 105-113.

Demian, P. and Balatsoukas, P. (2012). ”Information retrieval from civil engineering repositories: importance of context and granularity.” *J. Comput. Civ. Eng.*, 26(6), 727–740.

Ding, X., Liu, B., and Yu, P. S. (2008). “A holistic lexicon-based approach to opinion mining.” *Proc. 2008 Intl. Conf. on Web Search and Data Mining*, 231-240.

Egozi, O., Markovitch, S., and Gabrilovich, E. (2011). “Concept-based information retrieval using explicit semantic analysis.” *J. ACM Trans. Inform. Syst.*, 29(2), Article 8.

El-Diraby, T. (2012). “Domain ontology for construction knowledge.” *J. Constr. Eng. Manage.*, 139(7), 768-784.

El-Gohary, N., Liu, L., El-Rayes, K., and Lv, X. (2014). *ICT Project R27-132 Incorporating NEPA into IDOT and MPO Planning Processes*, Illinois Center for Transportation, Rantoul, IL.

Fan, H., and Li, H. (2013). “Retrieving similar cases for alternative dispute resolution in construction accidents using text mining techniques.” *Autom. In Constr*., 34, 85-91.

Fan, H., Xue, F., and Li, H. (2015). ”Project-based as-needed information retrieval from unstructured AEC documents.” *J. Manage. Eng*., 31(1), 1-12.

Fang, X., and Zhan, J. (2015). “Sentiment analysis using product review data.” *Journal of Big Data*, 2(1), 5.

Fernandez, M., Cantador, I., Lopez, V., Vallet, D., Castells, P., and Motta, E. (2011). “Semantically enhanced information retrieval: an ontology-based approach.” *J. Web. Semant.*, 9 (4), 434-452.

Florida Department of Transportation (FDOT). (2005). “Developing ETDM key performance measures.” *ETDM performance management plan*, FDOT, Tallahassee, FL, 12-21.

FDOT. (2006). “Chapter II: ETDM process.” *ETDM manual*, FDOT, Tallahassee, FL, 28-45.

Federal Highway Administration (FHWA). (2000). “Reasons for EIS project delays.” *Environmental review toolkit, accelerating project delivery*, <http://environment.fhwa.dot.gov/strmlng/eisdelay.asp> (September 2, 2013).

FHWA. (2002). “More than a philosophy: Maine's integrated transportation decision making process.” *Environmental review toolkit, accelerating project delivery*, <http://www.environment.fhwa.dot.gov/strmlng/newsletters/oct02nl.asp> (December 23, 2012).

FHWA. (2007a). “Colorado: strategic transportation, environmental, and planning process for urban places.” *Environmental review toolkit, planning and environmental linkages*, <http://environment.fhwa.dot.gov/integ/case_colorado.asp> (October 13, 2012).

FHWA. (2007b). “Florida: efficient transportation decision-making process.” *Environmental review toolkit, planning and environmental linkages*, <http://environment.fhwa.dot.gov/integ/case_florida.asp> (September 13, 2012).

FHWA. (2007c). “Indiana's streamlined EIS procedures.” *Environmental review toolkit, planning and environmental linkages*, <http://environment.fhwa.dot.gov/integ/case_indiana.asp> (November 13, 2012).

FHWA. (2007d). “Maine's integrated transportation decision-making (ITD) process.” *Environmental review toolkit, planning and environmental linkages*, <http://environment.fhwa.dot.gov/integ/case_maine.asp?\> (October 23, 2012).

FHWA. (2011). “Making a planning study viable for NEPA.” *Guidance on using corridor and subarea planning to inform NEPA*, FHWA, Washington, DC, 24-32.

FHWA. (2013). “National environmental streamlining initiatives.” *Environmental review toolkit, accelerating project delivery*, <http://environment.fhwa.dot.gov/strmlng/casestudies/index.asp> (May 3, 2013).

FHWA. (2016). “A Summary of Highway Provisions”, *Fixing America's Surface Transportation Act or “FAST Act”, <https://www.fhwa.dot.gov/fastact/summary.cfm>* (Nov. 2, 2017).

Fernández-Gavilanes, M., Álvarez-López, T., Juncal-Martínez, J., Costa-Montenegro, E. and González-Castaño, F.J. (2016). “Unsupervised method for sentiment analysis in online texts.” *Expert Syst. Appl.*, 58, 57-75.

Gamallo, P., and Garcia, M. (2014). “Citius: A naive-bayes strategy for sentiment analysis on english tweets.” *Proc. of SemEval*, 171-175.

García-Pablos, A., Cuadros, M., and Rigau, G. (2017). “W2VLDA: almost unsupervised system for aspect based sentiment analysis.” *Expert Systems with Applications*, 91, 127-137.

Ghani, R., Probst, K., Liu, Y., Krema, M., and Fano, A. (2006). “Text mining for product attribute extraction.” *ACM SIGKDD Explorations Newsletter 2006*, 8(1), 41-48.

Gong, Z., Cheang, C. W., and U, L. H. (2006). “Multi-term web query expansion using WordNet.” *Proc., 17th Int. Conf. Datab. and Exp. Syst. Appl.*, Springer, Berlin, Heidelberg, 379-388.

Goldberg, Y. (2016). “A Primer on Neural Network Models for Natural Language Processing.” *J. Artif. Intell. Res.*, 57, 345-420.

Griffiths, T. L., and Steyvers, M. (2004). “Finding scientific topics.” *Proc. of the National academy of Sciences*, 101(1), 5228-5235.

Gupta, P., Bali, K., Banchs, R. E., Choudhury, M., and Rosso, P. (2014). "Query expansion for mixed-script information retrieval." *Proc., 36th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, ACM, New York, NY, 677-686.

Hahm, G. J., Yi, M. Y., Lee, J. H., and Suh, H. W. (2014). "A personalized query expansion approach for engineering document retrieval." *ADV ENG INFORM*, 28(4), 344-359.

Hahm, G. J., Lee, J. H., and Suh, H. W. (2015). "Semantic relation based personalized ranking approach for engineering document retrieval." *ADV ENG INFORM*, 29(3), 366-379.

Han, J., Kamber, M., and Pei, J. (2011). *Data mining: concepts and techniques,* 3rd Ed., Morgan Kaufmann, Burlington, MA, 246-248.

Han, X., Wei, W., Miao, C., Mei, J. P., and Song, H. (2014). "Context-aware personal information retrieval from multiple social networks." *Computational Intelligence Magazine*, IEEE, 9(2), 18-28.

Honderich, T. (1995). *The Oxford companion to philosophy*, Oxford University Press, New York, NY.

Hsu, C. W., Chang, C. C., and Lin, C. J. (2003). *A practical guide to support vector classification*.

Hu, M., and Liu, B. (2004). "Mining opinion features in customer reviews." *Proc., National Conf. on Artificial Intelligence,* 755-760.

Hu, X., Tang, J., Gao, H., and Liu, H. (2013). "Unsupervised sentiment analysis with emotional signals." *Proc. 22nd international conference on World Wide Web*. 607-618.

Illinois Center for Transportation (ICT). (2014). "Guidelines developed to streamline NEPA and IDOT/MPO transportation planning processes." <http://ict.illinois.edu/2014/07/23/guidelines-developed-to-streamline-nepa-and-idotmpo-transportation-planning-processes-2/> (June 23, 2014).

Illinois Department of Transportation (IDOT). (2006). *MPO planning process: overview of the transportation planning process in urbanized areas*, IDOT, Springfield, IL.

IDOT. (2007). *Illinois state transportation plan special report: transportation funding*, IDOT, Springfield, IL.

IDOT. (2010). *Bureau of design and environment manual,* IDOT, Springfield, IL.

Indiana Department of Transportation (INDOT), and FHWA. (2007). *Streamlined environmental impact statement procedures*, Indiana Department of Transportation (INDOT), Indianapolis, IN.

Jakob, N., and Gurevych, I. (2010). "Extracting opinion targets in a single-and cross-domain setting with conditional random fields." *Proc., 2010 Conf. on Empirical Methods in NLP*, 1035-1045.

Jayatilaka, B., and Lee, J. (2003). "An epistemological taxonomy for knowledge management systems analysis." *Proc., 36th Hawaii Int. Conf. Syst. Sc.*, IEEE, Washington, D.C., 10-20.

Jiang, J. J., and Conrath, D. W. (1997). "Semantic similarity based on corpus statistics and lexical taxonomy." *Proc., Int. Conf. Research in Comput. Linguist.*, ACM, New York, NY.

Jiménez-Zafra, S. M., Martín-Valdivia, M. T., Martínez-Cámara, E., and Ureña-López, L. A. (2016). "Combining resources to improve unsupervised sentiment analysis at aspect-level." *Journal of Information Science,* 42(2), 213-229.

Jin, W., Ho, H. H., and Srihari, R.K. (2009). "A novel lexicalized HMM-based learning framework for web opinion mining." *Proc., 26th Ann. Intl. Conf. on Machine* Learning, 465-472.

Jin, W., Ho, H. H., and Srihari, R. K. (2009). "OpinionMiner: a novel machine learning system for web opinion mining and extraction." *Proc. 15th ACM SIGKDD*, 1195-1204.

Joachims, T. (2002). *Learning to classify text using support vector machines: Methods, theory and algorithms (Vol. 186),* Norwell: Kluwer Academic Publishers.

Joachims, T. (2008). "SVM^{hmm} Sequence Tagging with Structural Support Vector Machines." <https://www.cs.cornell.edu/people/tj/svm_light/svm_hmm.html> (Dec. 2, 2016).

Kara, S., Alan, O., Sabuncu, O., Akpinar, S., Cicekli, N., and Alpaslan, F. (2012). "An ontology-based retrieval system using semantic indexing." *J. Inform. Syst.*, 37(4), 294-305.

Katiyar, A., and Cardie, C. (2016). "Investigating LSTMs for Joint Extraction of Opinion Entities and Relations." *Proc., 54th Meeting of Assoc. for Comput. Linguist.*, 919-929.

Keck, D., Pate, H., Scolaro, A.J., Bloch, A., and Ryan, C. (2010). *National Cooperative Highway Research Program Report 662 Accelerating Transportation Project and Program Delivery: Conception to Completion*, TRB, Washington D.C, 16-29.

Khan, F. H., Bashir, S., and Qamar, U. (2014). "TOM: Twitter opinion mining framework using hybrid classification scheme". *Decis. Support Syst.*, 57, 245-257.

Kim, Y. (2014). "Convolutional neural networks for sentence classification." *arXiv preprint arXiv*:1408.5882.

Kiryakov, A., Popov, B., Terziev. I., Manov, D., and Ognyanoff, D. (2004)."Semantic annotation, indexing and retrieval." *J. Web. Semant.*, 2(1), 49-79.

Knight, S., Shum, S. B., and Littleton, K. (2013). "Tracking epistemic beliefs and sense making in collaborative information retrieval." *Proc., Computer Supported Coop. Work (CSCW) 2013*, ACM, New York, NY.

Korobov, M. (2015). "skearn-crfsuite." < https://sklearn-crfsuite.readthedocs.io/en/latest/> (Jan. 2, 2017).

Laerd Statistics. (2013). “The ultimate IBMSPSS guides.” <https://statistics.laerd.com/> (September. 18, 2015).

Lafferty, S. D. (2016). “IDOT still moving forward on Illiana toll road.” <http://www.chicagotribune.com/suburbs/daily-southtown/news/ct-sta-idot-pursues-illiana-st-1009-20161007-story.html> (Jan. 3, 2018).

Lafferty, J., McCallum, A., and Pereira, F. C. (2001). “Conditional random fields: Probabilistic models for segmenting and labeling sequence data.” *Proc., 18th Intl. Conf. on Machine Learning 2001 (ICML 2001)*, 282-289.

Lan, M., Tan, C. L., Su, J., & Lu, Y. (2009). “Supervised and traditional term weighting methods for automatic text categorization.” *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(4), 721-735.

Lavergne, T., Cappé, O., and Yvon, F. (2010). “Practical very large scale CRFs.” *Proc., 48th Ann. Meeting of Assoc. for Comput. Linguist.*, 504-513.

Leacock, C., and Chodorow, M. (1998). “Combining local context and WordNet similarity for word sense identification.” *WordNet: An Electronic Lexical Database*, MIT Press, Cambridge, MA, 265-283.

Levy, O., Goldberg, Y., and Dagan, I. (2015). “Improving distributional similarity with lessons learned from word embeddings.” *Transactions of the Association for Computational Linguistics*, 3, 211-225.

Li, F., Han, C., Huang, M., Zhu, X., Xia, Y. J., Zhang, S., and Yu, H. (2010). “Structure-aware review mining and summarization.” *Proc., 23rd Intl. Conf. on Comput. Linguist.*, 653-661.

Li, F., Pan, S. J., Jin, O., Yang, Q., and Zhu, X. (2012). “Cross-domain co-extraction of sentiment and topic lexicons.” *Proc., 50th Ann. Meeting of Assoc. for Comput. Linguist.*, 1, 410-419.

Li, Y., Bandar, Z. A., and Mclean, D. (2003). "An approach for measuring semantic similarity between words using multiple information sources." *IEEE Trans. Knowl. Data Eng.*, 15(4), 871-882.

Li, Y., and Bontcheva, K. (2007). "Hierarchical, perceptron like learning for ontology based information extraction." *Proc., 16th Int. World Wide Web Conf.*, ACM, New York, NY, 777-786.

Lin, C., and He, Y. (2009). "Joint sentiment/topic model for sentiment analysis." *Proc. 18th ACM Conf. on Info. and Knowl. Manage.*, 375-384.

Lin, D. (1998). "An information-theoretic definition of similarity." *Proc., 15th Int. Conf. Mach. Learn.*, ACM, New York, NY, 112-120.

Lin, H. T., and Chi, N. W., and Hsieh, S. H. (2012). "A concept-based information retrieval approach for engineering domain-specific technical documents." *Adv. Eng. Inform.*, 26, 349–360.

Lin, K. Y., and Soibelman, L. (2007). "Knowledge-assisted retrieval of online product information in architectural/engineering/construction." *J. Constr. Eng. Manage.*, 133, 871-879.

Lin, K. Y., and Soibelman, L. (2009). "Incorporating domain knowledge and information retrieval techniques to develop an architectural/engineering/construction online product search engine." *J. Comput. Civ. Eng.*, 23(4), 201-210.

Liu, S., McMahon, C. A., Darlington, M. J., Culley, S. J., and Wild, P. J. (2006). "A computational framework for retrieval of document fragments based on decomposition schemes in engineering information management." *ADV ENG INFORM*, 20(4), 401-413.

Liu, B. (2007). *Web data mining: Exploring hyperlinks, contents, and usage data*, Springer, Berlin, Heidelberg.

Liu, B. (2012). “Sentiment analysis and opinion mining.” *Synthesis lectures on human language technologies*, 5(1), 1-167

Liu, B., and Zhang, L. (2012). “A survey of opinion mining and sentiment analysis”. *Mining text data*, Springer, US, 415-463.

Liu, Y., and Zhang, M. (2018). *Neural Network Methods for Natural Language Processing*.

Lv, X., and El-Gohary, N. M. (2016a). “Discovering context-specific integration practices for integrating NEPA into statewide and metropolitan project planning processes.” *J. Constr. Eng. Manage*, 142(11), 04016056.

Lv, X., and El-Gohary, N. M. (2016b). “Semantic annotation for supporting context-aware information retrieval in the transportation project environmental review domain.” *J. Comput. Civ. Eng.*, 30(6), 04016033.

Lv, X., and El-Gohary, N. (2015). “Semantic annotation for context-aware information retrieval for supporting the environmental review of transportation projects.” *2015 ASCE International Conference on Computing in Civil Engineering (IWCCE)*, University of Texas at Austin, Austin, TX.

Lv, X., and El-Gohary, N.M. (2016). “Text analytics for supporting stakeholder opinion mining for large-scale highway.” *2016 International Conference on Sustainable Design, Engineering and Construction (ICSDEC)*, University of Arizona, Tempe, AZ.

Lv, X., and El-Gohary, N.M. (2016). “Semantic-based information retrieval for supporting project decision making.” *2016 International Conference on Computing in Civil and Building Engineering (ICCBE)*, Osaka University, Osaka, Japan.

Lv, X., and El-Gohary, N.M. (2016). “Enhanced context-based document relevance assessment and ranking for improved information retrieval to support environmental decision making” *Journal of Advanced Engineering Informatics*, 30(4), 737 - 750.

Lv, X., and El-Gohary, N.M. (2017). “Stakeholder opinion classification for supporting large-scale transportation project decision making.” 2017 *International Workshop on Computing for Civil Engineering (IWCCE),* University of Washington, Seattle, WA.

MacDonald, T., and Lidov, P. (2005). *STEP UP phase I report*, Colorado Department of Transportation (CDOT), Denver, CO, 12-54.

MacDonald, T., and Lidov, P. (2007). *STEP UP phase II statewide implementation report*, CDOT, Denver, CO, 23-34.

Mallett W.J., and Luther L. (2011). *Accelerating Highway and Transit Project Delivery: Issues and Options for Congress*, Congressional Research Service, Washington, D.C, 10-14.

Manning, C. D., Raghavan, P., and Shutze, H. (2009). *Introduction to information retrieval*, Cambridge University Press, Cambridge, England, 32-34.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., and McClosky, D. (2014). “The Stanford CoreNLP Natural Language Processing Toolkit.” In *Proc., 52nd Ann. Meeting of Assoc. for Comput. Linguist.: System Demonstrations*, 55-60.

Mao, W., and Chu, W. W. (2007). “The phrase-base vector space model for automatic retrieval of free-text medical documents.” *J. Data Knowl. Eng.*, 61(1), 76-92.

Marrese-Taylor, E., Velásquez, J. D., and Bravo-Marquez, F. (2014). “A novel deterministic approach for aspect-based opinion mining in tourism products reviews.” *Expert Systems with Applications*, 41(17), 7764-7775.

McCallum, A., Freitag, D., and Pereira, F. C. (2000). “Maximum Entropy Markov Models for Information Extraction and Segmentation.” *Proc., 17th Intl. Conf. on Machine Learning,* 591-598.

McGibbney, L. J., and Kumar, B. (2011). “A Knowledge-directed information retrieval and management framework for energy performance building regulations.” *Computing in Civil Engineering (2011),* ASCE, Reston, VA, 339-346.

Medhat, W., Hassan, A., and Korashy, H. (2014). “Sentiment analysis algorithms and applications: A survey.” *Ain Shams Eng. J.,* 5(4), 1093-1113.

Mei, Q., Ling, X., Wondra, M., Su, H., and Zhai, C. (2007). “Topic sentiment mixture: modeling facets and opinions in weblogs.” *Proc. 16th Intl. Conf. on World Wide Web*, 171-180.

Meng, L., Huang, R., and Gu, J. (2013). “A review of semantic similarity measures in wordnet.” *Int. J. Hybrid Inform. Tech.,* 6(1), 1-12.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). “Distributed representations of words and phrases and their compositionality.” *Adv. Neural Inform. Process. Syst.,* 3111–3119.

Montoyo, A., MartíNez-Barco, P., and Balahur, A. (2012). “Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments.” *Decis. Support Syst.,* 53(4), 675-679.

Muis, K. R. (2004). “Personal epistemology and mathematics: A critical review and synthesis of research.” *Rev. Edu. Research*, 74, 317-377.

Mukherjee, A., and Liu, B. (2012). “Aspect extraction through semi-supervised modeling.” *Proc., 50th Ann. Meeting of Assoc. for Comput. Linguist.*, 339-348.

Nesic, S., Crestani, F., Jazayeri, M., and Gasevic, D. (2010). “Concept based semantic annotation, indexing and retrieval of office-like document units.” *Proc., RIAO '10 Adaptively, Personalization and Fusion of Heterog. Inform.*, ACM, New York, NY, 134-135.

Ng, A. Y. (2004). “Feature selection, L1 vs. L2 regularization, and rotational invariance.” *Proc., 21st Intl. Conf. on Machine Learning,* 78 - 86.

Nik-Bakht, M., and El-Diraby, T. E. (2016). “Communities of interest-interest of communities: social and semantic analysis of communities in infrastructure discussion networks.” *Computer-Aided Civil and Infrastructure Engineering*, 31(1), 34-49.

Okazaki N. (2007). “CRFsuite: a fast implementation of Conditional Random Fields (CRFs).” < http://www.chokkan.org/software/crfsuite/> (Oct. 2, 2016).

Ozcan, R., and Aslandogan, Y. A. (2005). “Concept based information access.” *Int. Symp. Inform. Tech.: Coding and Comput.*, 1, 4-6.

Pang, B., Lillian, L., and Shivakumar, V. (2002).“Thumbs up?: sentiment classification using machine learning techniques.” *Proc. Conf. on Empirical Methods in NLP (EMNLP-2002).*

Pang, B., and Lee, L. (2008). “Opinion mining and sentiment analysis.” *Foundations and Trends in Information Retrieval*, 2(1–2), 1-135.

Pedregosa, F., et al. (2011). “Scikit-learn: Machine learning in python.” *J. Mach. Learn. Res.*, 12(10), 2825–2830.

Pennington, J., Socher, R., and Manning, C. D. (2014). “Glove: Global vectors for word representation.” *Proc., 2014 Conf. on Empirical Methods in NLP*, 14, 1532-1543.

Petrakis, E. G., Varelas, G., Hliaoutakis, A., and Raftopoulou, P. (2006). “Design and evaluation of semantic similarity measures for concepts stemming from the same or different ontologies.” *Proc., 4th Workshop on Multimedia Semant. (WMS'06),* 44-52.

Popescu, A. M., and Etzioni, O. (2007). “Extracting product features and opinions from reviews.” *Natural Language Processing and Text Mining*, 9-28.

Popov B., Kiryakov A., Kirlov A., Manov D., Ognyanoff D., and Goranov M. (2003). “KIM-semantic annotation platform.” *Semantic Web-ISWC 2003,* Springer, Berlin, Heidelberg, 834-849.

Poria, S., Cambria, E., Ku, L. W., Gui, C., and Gelbukh, A. (2014) “A rule-based approach to aspect extraction from product reviews.” *Proc., 2nd Workshop on NLP for Social Media (SocialNLP),* 28-37.

Poria, S., Cambria, E., and Gelbukh, A. (2016). “Aspect extraction for opinion mining with a deep convolutional neural network.” *Knowl.-Based Syst.,* 108, 42-49.

Poria, S., Chaturvedi, I., Cambria, E., and Bisio, F. (2016). “Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis.” *Proc. Intl. Joint Conf. in Neural Networks (IJCNN 2016),* 4465-4473.

Qiu, G., Liu, B., Bu, J., and Chen, C. (2011). “Opinion word expansion and target extraction through double propagation.” *Comput. Linguist.*, 37(1), 9-27.

Qu, L., Georgiana I., and Gerhard W. (2010) "The bag-of-opinions method for review rating prediction from sparse text patterns." *Proceedings of the 23rd Intl. Conf. on Computational Linguistics*. Association for Computational Linguistics, Stroudsburg, PA, 913-921.

Raghavan, H., and Iyer, R. (2007). “Evaluating vector-space and probabilistic models for query to ad matching.” *Information Retrieval in Advertising*.

Ravi, K., and Ravi, V. (2015). “A survey on opinion mining and sentiment analysis: tasks, approaches and applications.” *Knowl.-Based Syst.*, 89, 14-46.

Rehurek, R., and Sojka, P. (2010). “Software framework for topic modelling with large corpora.” *Proc. of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45-50.

Resnik, P. (1995). “Using information content to evaluate semantic similarity in a taxonomy.” *Proc., 14th Int. Joint Conf. Artif. Intell.*, ACM, New York, NY, 1, 448-453.

Rey, D., and Neuhäuser, M. (2011). “Wilcoxon-signed-rank test.” *International encyclopedia of statistical science*, Springer, Heidelberg, Germany, 1658-1659.

Roelleke, T. (2013). *Information Retrieval Models: Foundations and Relationships*, Synthesis Lectures on Information Concepts, Retrieval, and Services, Morgan & Claypool publishers, Toronto, Canada, 1-163.

Sayeedunnissa, S. F., Hussain, A. R., and Hameed, M. A. (2013). “Supervised opinion mining of social network data using a bag-of-words approach on the cloud.” *Proc. 17th Intl Conf. BIC-TA*, 299-309.

Shariaty, S., and Moghaddam, S. (2011). “Fine-grained opinion mining using conditional random fields.” *Data Mining Workshops (ICDMW), 2011 IEEE 11th Intl. Conf. on 2011*, 109-114.

Sharma, A., and Dey, S. (2012). “A comparative study of feature selection and machine learning techniques for sentiment analysis.” *Proc. 2012 ACM research in applied computation symposium*, 1-7.

Shu, L., Xu, H., and Liu, B. (2017). “Lifelong Learning CRF for Supervised Aspect Extraction.” *arXiv*: 1705.00251.

Simperl, E. P. B., and Mochol, M. (2006). “Cost estimation for ontology development.” *Proc. 9th Business Information System*, 4-16

Singhal, A. (2001). “Modern information retrieval: A brief overview.” *IEEE Data Eng. Bull.*, 24(4), 35-43.

Slimani, T. (2013). "Description and evaluation of semantic similarity measures approaches." *Int. J. Comput. Appl.*, 80 (10), 25-33.

Soibelman, L., Wu, J., Caldas, C., Brilakis, I., and Lin, K. Y. (2007). "Management and analysis of unstructured construction data types." *ADV ENG INFORM*, 22(1). 15-27.

Spy Pond Parteners, LLC, University of Minnesota Center for Transportation Studies, and Tucker, S. (2009). *NCHRP report 643: implementing transportation knowledge networks*, TRB, Washington D.C., 3-9.

Steup, M. (2011). "Epistemology", *Stanford encyclopedia of philosophy (*winter *2011 ed.)*, <http://plato.stanford.edu/archives/win2011/entries/epistemology/> (August 18, 2014).

Stevenson, M., and Greenwood, M. A. (2005). "A semantic approach to IE pattern induction." *Proc., 43rd Ann. Meet. Assoc. for Comput. Linguist.*, ACM, New York, NY, 379-386.

Sutton, C., and McCallum, A. (2012). "An introduction to conditional random fields." *Foundations and Trends in Machine Learning,* 4(4), 267-373.

Szymański, P., and Kajdanowicz, T. (2017). "A scikit-based Python environment for performing multi-label classification." *arXiv*: 1702.01460.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. (2011). "Lexicon-based methods for sentiment analysis." *Comput. Linguist.,* 37(2), 267-307.

Tan, P., Steinbach, M., and Kumar, V. (2013). *Introduction to data mining*, Addison-Wesley, Boston, MA, 329-311.

Toh, Z., and Wang, W. (2014). "Dlirec: Aspect term extraction and term polarity classification system." *Proc. 8th Intl. Workshop on Semantic Evaluation (SemEval 2014)*, 235-240.

Transportation Research Board (TRB). (2014). "NCHRP 20-97." *Improving findability and relevance of transportation information,*

<http://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=3665> (August 23, 2014).

Tsoumakas, G., and Katakis, I. (2007). “Multilabel classification: An overview.” *Int. J. Data Warehouse. Min.*, 3(3), 1-13.

Turney, P. D., and Pantel, P. (2010). “From frequency to meaning: Vector space models of semantics.” *Journal of artificial intelligence research*, 37, 141-188.

Ur-Rahman, N., and Harding, J. A. (2012). “Textual data mining for industrial knowledge management and text classification: A business oriented approach.” *Expert Syst. Appl.*, 39(5), 4729-4739.

U.S. Government Printing Office (USGPO). (1998). “Section 1309- environmental streaming.” *Transportation Equity Act for the 21st Century*, USGPO, Washington, DC, 132-145.

USGPO. (2007). “Section 6002-efficient environmental review for project decision-making.” *Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users*, USGPO, Washington, DC, 12-46.

USGPO. (2013). “Section 1301-1323: accelerating project delivery.” *Moving Ahead for Progress in the 21st Century*, USGPO, Washington, DC, 78-95.

Venner Consulting, Institute for Natural Resources, Oregon State University, and Parametrix, Inc. (2012). *Expedited planning and environmental review of highway projects*, TRB, Washington, D.C., 13-14.

Walters, W. H. (2011). “Comparative recall and precision of simple and expert searches in Google Scholar and eight other databases.” *Portal: Libraries and the Academy*, 11(4), 971-1006.

Wang, C., and Akella, R. (2015). “Concept-based relevance models for medical and semantic information retrieval.” *Proc., 24th ACM Intl. Conf. Information and Knowl. Manage*., ACM, New York, NY, 173-182.

Williams, T. P., and Gong, J. (2014). “Predicting construction cost overruns using text mining, numerical data and ensemble classifiers.” *Automation in Construction*, 43, 23-29.

Wu, Z., and Palmer, M. (1994). “Verb semantics and lexical selection.” *Proc., 32nd Ann. Meet. Assoc. for Comput. Linguist.*, ACM, New York, NY, 133-128.

Yang, B., and Cardie, C. (2012). “Extracting opinion expressions with semi-Markov conditional random fields.” *Proc. 2012 Joint Conf. on Empirical Methods in NLP and Comput. NLP*, 1335-1345.

Yang, B., and Cardie, C. (2014). “Context-aware Learning for Sentence-level Sentiment Analysis with Posterior Regularization.” *Proc. 52nd Ann. Meeting of Assoc. for Comput. Linguist.*, 325-335.

Yu, W., and Hsu, J. Y. (2013). “Content-based text mining technique for retrieval of CAD documents.” *Autom. in Constr.*, 31, 65-74.

Yu, J., Zha, Z. J., Wang, M., and Chua, T. S. (2011). “Aspect ranking: identifying important product aspects from online consumer reviews.” *Proc. 49th Ann. Meeting of Assoc. for Comput. Linguist.: Human Language Technologies*, 1, 1496-1505.

Zhai, C. (2008). *Statistical language models for information retrieval*, Synthesis Lectures on Human Language Technologies, Morgan & Claypool publishers, Toronto, Canada, 1-141.

Zhai, Z., Liu, B., Xu, H., and Jia, P. (2011). "Clustering product features for opinion mining." *Proceedings of the fourth ACM Intl. Conf. on Web search and data mining*. ACM, New York, NY, 347-354.

Zhang, L., and El-Gohary, N. M. (2015). "Epistemology-based context-aware semantic model for sustainable construction practices." *J. Constr. Eng. Manage.*, 142(3), 04015084.

Zhang, L., and Liu, B. (2014). "Aspect and entity extraction for opinion mining." *Data mining and knowledge discovery for big data*, Springer, Berlin, Heidelberg, 1-40.

Zhang, M. L., and Zhou, Z. H. (2007). "ML-KNN: A lazy learning approach to multilabel learning." *Pattern Recognit.*, 40(7), 2038-2048.

Zhang, X., Jing, L., Hu, X., Ng, M., and Zhou, X. (2007). "A comparative study of ontology-based term similarity measures on PubMed document clustering." *Proc., 12th Int. Conf. Datab. Syst. for Adv. Appl.*, Springer, Berlin, Heidelberg, 115-126.

Zainuddin, N., and Selamat, A. (2014). "Sentiment analysis using support vector machine." *Proc. 2014 International Conference on Computer, Communications, and Control Technology*, 333-337.

Zou, H., and Hastie, T. (2005). "Regularization and variable selection via the elastic net." *J. Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301-320.