Intelligent Construction Case Study Illustration System Using Natural Language Processing and Image Searching

Jiansong Zhang, jiansong.zhang@wmich.edu
Western Michigan University, United States

Abiola Akanmu, abiola.akanmu@wmich.edu
Western Michigan University, United States

Abstract

With the rapid advancement in science and engineering disciplines, the amount of knowledge exploded. Therefore, it is more and more difficult for a single person to become an expert in multiple domains. In this regard, an intelligent illustration system is beneficial to help people quickly understand textual description of case studies in different domains. This is especially the case in construction engineering and management education, where the students may come from different backgrounds (e.g., architecture, civil engineering and non-civil related domains). And the students may not understand many terminologies used in construction projects (e.g., toenailing, sandblasting and Canadian snow load) which could hinder them from adequately analyzing construction case studies. To address this challenge, this paper proposes a new intelligent illustration system which takes a textual description of a construction case study and renders a set of structured images with each image illustrating an important concept in the textual description. The system is based on natural language processing and keywords-based image searching techniques. Part-of-speech (POS) tagging and dependency parsing are used to process a sentence to generate POS tags for each word, and dependency relations, respectively. Both POS tags and dependency relations are used in extracting the interesting concepts and relations from the case study. The concepts are then fed into the image searching process as the keywords. The retrieved images are organized into structured images to illustrate the interesting concepts, where the image result for each concept is connected to another image result using the corresponding relation. A preliminary implementation of the system was tested on a construction management case study. All interesting concepts and relations were successfully extracted and automatically assigned result images for illustrations. It was observed that some result images were not quite relevant to the input case study. As such, this will be addressed in future work by adding contextual terms to image search keywords.

Keywords: Knowledge modeling, construction education, natural language processing, image retrieval, intelligent systems

1 Introduction

Case studies are widely used in engineering education to promote active learning, boost positive attitudes towards learning, and enhance students' ability to deal with open-ended problems. With the increasing efforts to bring industry practice to classrooms, case studies can help convey the complexity and ambiguity of the practical world while also providing realism to students' learning (Newson & Norbert 2011). In the context of civil and construction engineering education, case studies are typically used to teach courses involving legal aspects of construction (Thomas 2009). In such courses, legal cases are written in narrative forms to illustrate the complex relationships and interactions between stakeholders and resources involved in issues related to differing site conditions, defective specifications, changes or misrepresentations. The case studies also describe the facts and opinions leading to decisions made by the involved parties. In spite of the advantages and potential of using case studies in engineering education, it has been reported (Rawlings et al 2014) that there are certain barriers to learning from case studies, especially when the learners come from different culture, backgrounds and disciplines (Rawlings et al 2014). One of such barriers includes the difficulty in understanding certain concepts and terminologies in the case description. This is further complicated by the inherent ambiguity in natural language text description.

Therefore, the objective of this research is to explore ways of transforming stodgy and difficult course topics into a more engaging and easy-to-understand learning experience. It is increasingly being recognized that visualization of concepts and terminologies can enhance students' cognitive process (Shirazi & Amir 2014). Thus, the authors propose an intelligent construction case study illustration system which is intended to illustrate concepts and relations in a case study textual description using structured images. The system leverages natural language processing (NLP) techniques and online image searching techniques. The following sections describe the background of these techniques, a detailed illustration of the intelligent construction case study illustration system, preliminary experiments on using the system, and experimental results.

2 Background

2.1 Natural language processing

Natural language processing (NLP) is a subdomain of artificial intelligence (AI) that targets enabling computers to mimic the natural language (text and speech) understanding and processing capabilities of humans (Cherpas 1992). Between speech processing and text processing, speech processing addresses the processing of spoken languages whereas text processing addresses the processing of written natural language texts. Text processing can be categorized into the following five levels: morphology, syntax, semantics, pragmatics, and discourse (Jurafsky & Martin 2006). Morphology studies the composition and structure of words (Carstairs-McCarthy 2002). Syntax studies the combination of words (Matthews 2002). Semantics studies the meaning of texts (Weekley 1917). Pragmatics studies the psychological, biological, and sociological phenomena in texts (Morris 1938). Discourse studies the way texts are interpreted (Sparks & Rapp 2010). Higher level text processing typically leverages output from lower level text processing, which leads to a pipeline-style methodology in most NLP efforts.

2.2 Part-of-speech tagging and dependency parsing

Part-of-speech (POS) tagging aims to tag each word in a sentence with their functional and lexical categories showing inherent language structures. Singular nouns (tag NN), adjectives (tag JJ), verbs (tag VB), and prepositions (tag IN) are examples of such POS tags (Galasso 2002). POS tagging is a syntax level NLP task. POS tags are widely used in various types of NLP tasks such as dependency parsing (syntax level task) (Klein & Manning 2004), information extraction (semantic level task), and semantic parsing (semantic level task). Dependency parsing parses a sentence into a representation with dependency relations between individual words (Marneffe et al 2006). Nominal subject (tag nsubj), direct object (tag dobj), and relative clause modifier (rcmod) are examples of such dependency relations (Marneffe et al 2006). In each dependency relation that connects two words, one of the words is named a "head" and the other word is named a "dependent." A major alternative of dependency parsing is phrase structure parsing, which parses a sentence into a representation with a "nesting of multi-word constituents" (Marneffe et al 2006). In comparison with phrase structure parsing, dependency parsing is less expressive, but simpler (Covington 2001). As a syntax level task, dependency parsing produces dependency relations that can be used by higher level NLP tasks such as information extraction and semantic parsing. Information extraction targets extracting needed information from (typically a large amount of) texts according to a predefined information template. Semantic parsing parses text into formal meaning representations (Poon & Domingos 2009).

2.3 NLP in construction research

In the construction research community, the use of NLP was observed since 2000s. Abuzir and Abuzir (2002) pioneered the exploration of using lexical syntactic features together with other structured information [e.g., HyperText Markup Language (HTML) tags] to extract civil engineering terms and relations from HTML documents. Caldas and Soibelman (2003) pioneered the exploration of machine learning-based text classification on construction documents. More recent works can be found around the areas of construction contract analysis (Al Qady & Kandil 2010), project document management (Al Qady & Kandil 2014), project information retrieval (Fan et al 2014), construction text classification (Salama & El-Gohary 2013; Zhou & El-Gohary 2015; Mart´ınez-Rojas et al 2013), and regulatory/specifications compliance checking (Zhang & El-Gohary 2015a, b; Li & Cai 2015). As far as the authors are concerned, no leverage of NLP in construction education has been reported.

2.4 Case study in construction education and training

The case study methodology which originates from social science, refers to an in-depth description and analysis of a single entity or phenomenon. The case study methodology can be descriptive or explanatory (Yin 1984; Schoenborn 2012). It is widely used in construction education and training. Construction education in the U.S. has seen a shift from the focus of construction science to more management based content (Burt et al 2008). The use of case study methodology well suits the description and analysis of management-based content. For example, Singh and Sakamoto (2001) analyzed a hypothetical case study in the setting of construction education, in regard of the liabilities and responsibilities of different parties such as owner, architect, contractor, subcontractor, and manufacturer. Grosskopf (2005) developed and used a case study to deliver and reinforce key business and financial competencies in construction students. Jaruhar (2008) used a case study model for educational simulation exercise development, for testing the usability of a simulation application for virtual construction scheduling and visualization. It has been found that using case studies can increase students' satisfaction of the learning experience and outcome (Grosskopf 2005). One challenge in using case study for construction education, nevertheless, is to efficiently convey the contextual information of a case to students. Because a construction case study typically covers multiple domain concepts, students with limited or no site experience may find it difficult to understand the domain concepts (especially those expressed using jargons), such as toenailing, sandblasting and Canadian snow load.

3 Proposed intelligent case study illustration system

"A picture is worth a thousand words." The authors propose to illustrate concepts in a case study using corresponding images found by online searching, where each image is found by image searching in online search engines using the name of the concept as a key word. The structure of the proposed system is shown in Figure 1. There are mainly seven processes in the system: (1) sentence splitting; (2) part-of-speech (POS) tagging, (3) concept extraction, (4) dependency parsing, (5) concept and relation extraction, (6) image searching, and (7) image result composition. The system takes a case study textual description as the input and generates structured images which illustrate the important concepts in the case study.

3.1 Sentence splitting

Sentence splitting is the process of splitting a piece of text into individual sentences, based on typical sentence boundaries such as periods, exclamation marks, and question marks. It is an intuitive task for humans but not so much for the machine. Abbreviations and proper nouns can easily confuse a sentence splitter if sufficient care is not taken, because it is common for abbreviations and proper nouns to contain periods which are classical sentence boundary indicators. Although perfect results are difficult to achieve and are highly context and language dependent, simple rule based sentence splitting methods can achieve very high performance (e.g., an accuracy of 99.1% to 99.5%) (Palmer 2000). Sentence splitting is needed because further NLP-based processing tasks are performed on the sentence basis.

3.2 Part-of-speech tagging

Part-of-speech (POS) tagging is the process of assigning each word in a sentence their lexical and functional categories in the form of POS tags. Also, numbers, punctuations, and symbols are also assigned tags. Examples of POS tags are RB (adverb), JJ (adjectives), NN (singular or mass noun), and VBD (past tense verb) (Santorini 1990). POS tags are the most widely used syntactic level information in many NLP tasks such as information extraction (Zhang & El-Gohary 2015a). In spite of the complexity of the POS tagging, higher than 97% token-level accuracy can be achieved using state-of-the-art POS taggers (Manning 2011). In the proposed system, POS tagging is used to help generate POS tags that are essential for the concept extraction process.

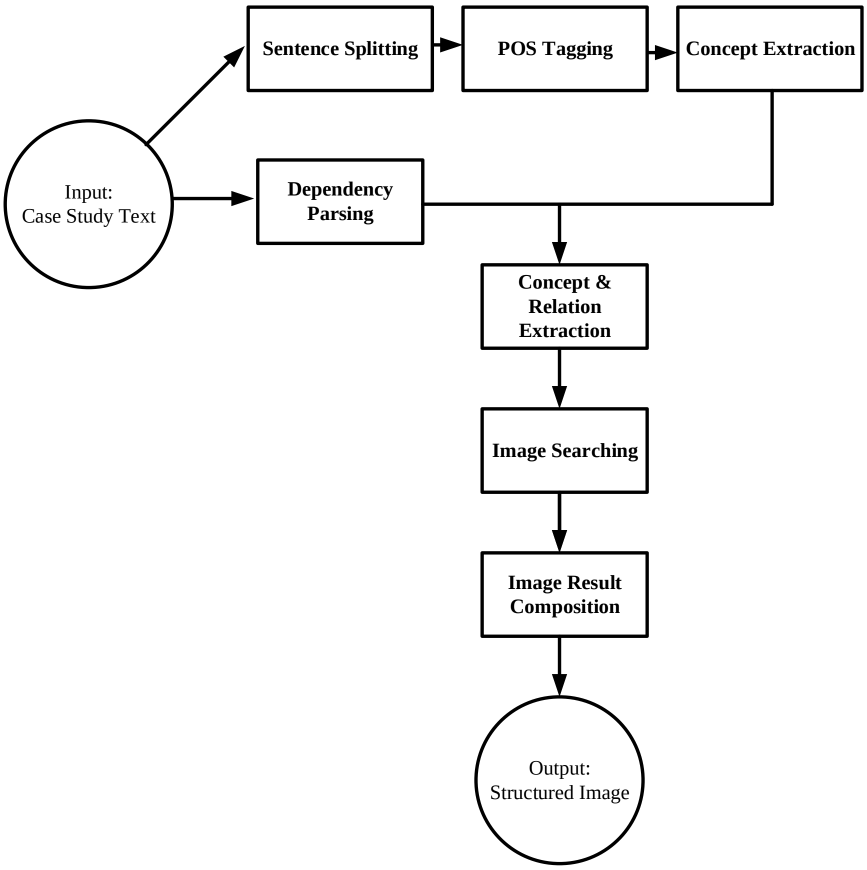

Figure 1 Proposed intelligent case study illustration system

3.3 Concept extraction

Concept extraction is the process of extracting all concepts from the case study. This step utilizes a pattern matching-based approach to identify all noun and noun phrases that represent concepts. The patterns are composed of sequences of POS tags. For example, the concept “low bidder” can be extracted based on its POS pattern “JJ NN.”

3.4 Dependency parsing

Dependency parsing is the process of parsing a sentence using dependency grammars to produce dependency relations. Together with phrase structure grammar, dependency grammar is one of the most widely used grammars for natural language parsing. However, parsing results using phrase structure grammar and dependency grammar are quite different. Parsing with phrase structure grammar produces a tree-like structure with “Sentence” as the root node, phrasal tags and POS tags as branches, and word tokens as leaves. Parsing with dependency grammar, on the other hand, produces a series of dependency relations. Each dependency relation is a pairwise relation between two words: one is called "head" and the other is called "dependent." For example, a pairwise relation amod (adjective modifier) is produced between "low" and "bidder" when "low bidder" is parsed using dependency grammar.

3.5 Concept and relation extraction

Concept and relation extraction is the process of extracting the concepts that are connected using interesting dependency relations. This is important to save computing power of the system from processing all concepts and all dependency relations. Many dependency relations are not interesting in our analysis, such as the det (determiner) relation which relates a determiner (i.e., a, an, the) to its head noun.

3.6 Image searching

Image searching is the process of searching for images (from the internet) that illustrate the selected concepts. The image searching process takes the selected concepts from the previous step and feeds them as key words to an image search engine.

3.7 Image result composition

In the image result composition stage, the found images are organized to form a structured result image by connecting the images with their dependency relation.

4 Preliminary experiment

To test the feasibility of the proposed intelligent case study illustration system, a prototype was built according to the system design. Because of the availability of vast amount of NLP resources in Python programming language, Python (Python v2.7.3) was used to build the prototype. This section describes the development and implementation of the proposed prototype system on a construction management case study.

4.1 Sentence splitting

Sentence splitting is a preprocessing task in many NLP applications. Writing a new sentence splitter is possible using regular expression based-rules in Python. The benefit of this is that the resulting sentence splitter can be tailored to a specific type of text and use case and therefore potentially achieve an accuracy closer to 100% than other general purpose sentence splitters. However, there are open-source sentence splitters with high performance. As described in Section 3.1, simple rule based methods can achieve an accuracy of higher than 99%. These open-source sentence splitters suite the fast prototyping purpose (discussed in this paper) with a reasonable performance. The built-in sentence splitter in the natural language toolkit (NLTK) (Bird et al 2009) of Python was used: PunkSentenceTokenizer in NLTK. The PunkSentenceTokenizer uses regular expressions to describe common patterns of punctuation usage and relies on pattern-based abbreviation detection to avoid the confusion caused by abbreviations (Russell 2011).

4.2 POS tagging and dependency parsing

To ensure the consistency between the tagging and parsing results, Stanford Parser was used for the POS tagging and dependency parsing. The POS tagger is a java implementation of the log-linear POS tagger (Toutanova et al 2003). The dependency parser is a neural network-based dependency parser (Chen & Manning 2014). Because the tools in Stanford parser are all Java based, and the main platform for the prototype system is Python, Jython (Jython 2.5.2) was used to enable the bi-directional communication.

4.3 Concept extraction

The concept extraction rules developed in Zhang & El-Gohary (2016) were used. There were 39 concept extraction rules and each rule uses one flattened POS pattern. Flattened POS patterns are patterns that include only terminal symbols (i.e., POS tags) and no non-terminal symbols (e.g., phrasal tags). Example POS patterns used, their meanings, and corresponding extraction examples are shown in Table 1. As shown in the table, each POS pattern takes one or more POS tags. The POS patterns consist of nouns with or without modifiers. Such POS patterns can be used to extract different types of concepts in the construction domain regardless of their meanings. For the details of the concept extraction method and the 39 rules, the readers are referred to Zhang & El-Gohary (2016).

Table 1 Example POS patterns with their meanings and extracted examples

POS Pattern	Meaning	Extraction Example
VBG	Gerund or present participle verb	Plastering
NN	Singular or mass noun	Contractor
NNS	Plural noun	Materials
JJ NNS	Adjective + plural noun	General conditions
NN NN	Two singular or mass noun	Metal deck
NN NN NN	Three consecutive singular or mass nouns	Metal flute deck
VBN NN NN	Past participle verb + two singular or mass noun	Approved shop drawing

4.4 Concept and relation extraction

An algorithm for concept and relation extraction was implemented in Python. Each concept and relation combination is represented as a two or three tuple: the result tuple. The algorithm for concept and relation extraction considers three types of interesting dependency relations. The **first** type of dependency relation considered is "nsubj," which indicates the relation between the subject (i.e., the dependent, a noun phrase as nominal subject) of a clause and its corresponding governor (i.e., the head) which can be a verb, adjective, or noun (Marneffe & Manning 2015). In this type of dependency relation, there is a distinction between three subtypes depending on whether there is a copula (represented in a cop dependency relation) or a direct object (represented in a dobj dependency relation): (1) if there is a copula, the three elements of the result tuple are tokens that correspond to the dependent of the nsubj relation, the copula, and the head of the nsubj relation; (2) if there is a dobj, the three elements of the result tuple are tokens that correspond to the head of the nsubj relation, the dependent of the nsubj relation, and the dependent of the dobj relation; (3) if there are no copula or dobj relations related to the nsubj relation, then the two elements of the result tuple are tokens that correspond to the head of the nsubj relation and the dependent of the nsubj relation. The **second** type of dependency relation considered is "prep," which indicates the relationship between the prepositional modifier (i.e., a prepositional phrase) and its corresponding head (prepositional phrase that serves to modify the meaning of the verb, adjective, noun, or even another preposition) (Marneffe & Manning 2015). In this type of dependency relation, only one subtype is considered: a direct object exists for the preposition (reflected by a "pobj" relation). In the result three tuple, the three elements correspond to the head of the "prep" relation, the dependent of the "prep" relation, and the direct object. The **third** type of dependency relation considered is "conj," which indicates the coordinating conjunction relation (e.g., "and," "or") between two elements (Marneffe & Manning 2015). In the result three tuple, the three elements correspond to the head of the "conj" relation, the coordinating conjunction relation type (e.g., "and" or "or"), and dependent of the "conj" relation. The three types of dependency relations considered and example result tuples for each type of dependency relation are shown in Table 2.

Table 2 Dependency relations and their example tuples

Dependency Relation	Tuple Element 1	Tuple Element 2	Tuple Element 3
Nsubj	Contractor	Claims	-
	Drawings	Showed	Narrow flute spaces
	Flute openings	Were	Wider
Prep	Installation	Of	Zonolite spray insulation

	Science area	Of	Building
	Applied	Over	Entire surface
Conj	Lathing	And	Plastering work
	Narrow flute spaces	And	Broad flute spaces

4.5 Image searching and image result composition

Image searching was conducted using Google image search. Beautiful soup Python library (Richardson 2015) was used to parse the resulting webpages from Google image search and pull out the image data. Image search was conducted on each tuple generated from the concept and relation extraction process. Among the three (or two) elements of each tuple, the first and third elements were used as keywords to find a corresponding image for both elements. Only the first returned image was used for each keyword. The two images were then connected using the second element of the tuple. In this way the output image for the tuple was generated. If the tuple consists of two elements only, then the image for the third element was replaced by a white background. Figure 2 shows the image composition result of the tuple <materials, and, labor>.

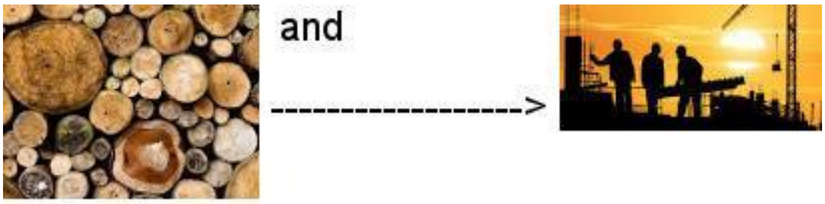

Figure 2 Image composition result for <materials, and, labor>

4.6 Case study use

The developed prototype system was tested on a case study used in the Contracts, Codes and Specifications class at Western Michigan University. The case study is focused on notice requirements (Thomas & Ellis 2007). The title and main body of the case study consists of 812 words. The case study has 6 paragraphs.

5 Experimental results and analysis

The prototype system extracted 214 concepts from the case study, and 190 concept-and-relation tuples from the case study. Each concept and relation tuple was automatically assigned a result image successfully. For example, Figure 2 shows the result image for the tuple <materials, and, labor>, Figure 3 shows the result image for the tuple <contractor, claims>, and Figure 4 shows the result image for the tuple <drawing, showed, narrow flute spaces>. However, there is a big variation in the result images regarding their relevancy in helping explain the case study. For example, Figure 2 shows a good illustration of materials and labors, and Figure 3 shows a good illustration of contractor. Nevertheless, the image corresponding to “drawing” is not really relevant in Figure 4. This is because when image search is conducted in a general purpose search engine, there is no context information provided regarding the topic or domain. Therefore, an English concept that can have different meanings in different domains or topics is not guaranteed to find a relevant image for illustrating the concept in the case study, especially because only the first returned image is used from each search.

Figure 3 Image composition result for <contractor, claims>

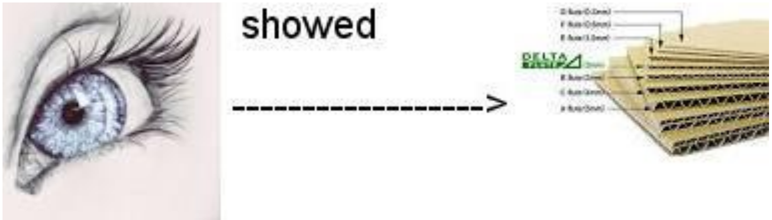

Figure 4 Image composition result for <drawing, showed, narrow flute spaces>

If topic or domain information can be added to keywords used for image searching, it may help improve the results. For example, when the term “construction” was added to the keyword “drawing,” which led to the result image shown in Figure 5 for the tuple <construction drawing, showed, narrow flute spaces>, the result was much better than that in Figure 4, to help illustrate the case study.

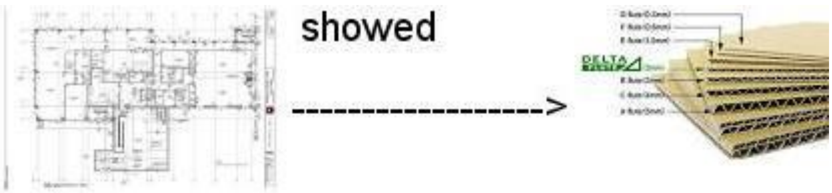

Figure 5 Image composition result for <construction drawing, showed, narrow flute spaces>

Similarly, Figure 6 shows the result image for <lathing, and, plastering work>, whereas Figure 7 shows the result image for <construction lathing, and, plastering work>, it can be seen that adding “construction” to “lathing” made the found image more relevant to the case study context.

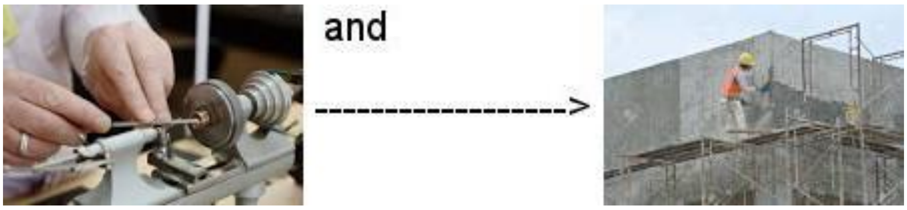

Figure 6 Image composition result for <lathing, and, plastering work>

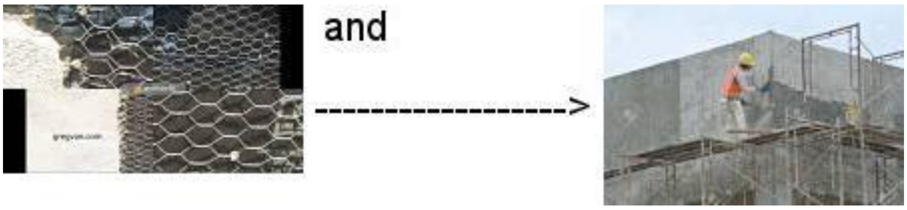

Figure 7 Image composition result for <construction lathing, and, plastering work>

6 Conclusion and future work

In this paper, the authors proposed an intelligent construction case study illustration system that takes a case study text, automatically analyzes the text for interesting concepts and relations, and generates image illustrations of the interesting concepts and relations. The system leverages natural language processing techniques to find part-of-speech (POS) tags of each word, and dependency relations in the text being analyzed. Both POS tags and dependency relations are used in the extraction of interesting concepts and relations. The interesting concepts and relations are then used to conduct image search in a search engine and the found images are used to compose the image result for illustrating the concepts and relations. A preliminary experiment conducted on a construction management case study focused on the notice requirement showed the proposed system successfully extracted the interesting concepts and relations, and generated image illustration for all concepts and relations. A key limitation of the proposed system is that some of the extracted images are not so relevant to the case study. By adding terms that describe the case study to the concept used for image searching, results can be improved. In future work, the authors will explore an effective and automated approach to incorporate context terms into the image searching keywords.

References

Abuzir, Y., & Abuzir, M. O. (2002). Constructing the civil engineering thesaurus (CET) using ThesWB. *Proc., Int. Workshop on Information Technology in Civil Engineering 2002*, ASCE, Washington, D.C., United States, November 2nd – 3rd, 2002. pp. 400-412.

Al Qady, M. & Kandil, A. (2010). Concept relation extraction from construction documents using natural language processing. *J. Constr. Eng. Manage.* 10.1061/(ASCE)CO.1943-7862.0000131. pp. 294-302.

Al Qady, M. & Kandil, A. (2014). Automatic classification of project documents on the basis of text content. *J. Comput. Civ. Eng.* 10.1061/(ASCE)CP.1943-5487.0000338, 04014043.

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python. O'Reilly Media. Sebastopol, CA, United States.

Burt, R., Hatipkarasulu, Y., & Nobe, M. E. C. (2008). Evolution of construction education in the United States: Two case studies. *ASC Proc., 44th Annual Conf.* Auburn University, Alabama, United States.

Caldas, C.H., & Soibelman, L. (2003). Automating hierarchical document classification for construction management information systems. *Autom. Constr.* 12(2003). pp. 395-406.

Carstairs-McCarthy, A. (2002). An introduction to English morphology: words and their structure. Ch 3. A word and its parts: roots, affixes and their shapes. pp. 16-16. <http://logic.sysu.edu.cn/ebookfull/UploadFiles_7160/200905/2009050617095579.pdf> (May 6th, 2016).

Chen, D., & Manning, C. (2014). A fast and accurate dependency parser using neural networks. *Proc., EMNLP 2014*. SIGDAT, Doha, Qatar, October 25th – 29th, 2014.

Cherpas, C. (1992). Natural language processing, pragmatics, and verbal behavior. *Anal. Ver. Behav.* 10. pp. 135-147.

Covington, M.A. (2001). A fundamental algorithm for dependency parsing. *Proc., 39th Annual ACM Southeast Conf*. University of Georgia, Athens, Georgia, March 16th – 17th, 2001. pp. 95-102.

Fan, H., Xue, F., & Li, H. (2014). Project-based as-needed information retrieval from unstructured AEC documents. *J. Manage. Eng.* 10.1061/(ASCE)ME.1943-5479.0000341, A4014012.

Galasso, J. (2002). Analyzing English grammar: an introduction to feature theory: a companion handbook. California State University Northridge, Northridge, CA, United States.

Grosskopf, K. (2005). Project simulation in construction finance education. *2005 AACE International Transactions*. DEV.02.

Jaruhar, S. (2008). Development of interactive simulations for construction engineering education. M.Sc. Thesis. The Pennsylvania State University, State College, PA, United States.

Jurafsky, D., & Martin, J.H. (2006). Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition. <http://stp.lingfil.uu.se/~santinim/ml/2014/JurafskyMartinSpeechAndLanguageProcessing2ed_draft%202007.pdf> (May 6th, 2016).

Jython 2.5.2. [Computer software]. <http://www.jython.org/> (May 04th, 2016).

Klein, D. & Manning, C.D. (2004). Corpus-based induction of syntactic structure: models of dependency and constituency. *Proc., Association for Computational Linguistics (ACL) 2004.* ACL, Forum Convention Centre Barcelona, Barcelona, Spain, July 21st – 26th, 2004.

Li, S. & Cai, H. (2015). Automated underground utility mapping and compliance checking using NLP-aided spatial reasoning. *Proc., Computing in Civil Engineering 2015*. University of Texas at Austin, Austin, TX, United States, Jun 21st – 23rd, 2015. pp. 419-426.

Manning, C.D. (2011). Part-of-speech tagging from 97% to 100%: is it time for some linguistics? *Proc., 12th International Conference on Intelligent Text Processing and Computational Linguistics*. CICLing, Tokyo, Japan, February 20th – 26th, 2011.

Marneffe, M. & Manning, C.D. (2015). Stanford typed dependencies manual. < http://nlp.stanford.edu/software/dependencies_manual.pdf> (Jun 21st, 2016).

Marneffe, M., MacCartney, B., & Manning, C.D. (2006). Generating typed dependency parses from phrase structure parses. *Proc., 5th Edition of the International Conference on Language Resources and Evaluation (LREC 2006)*. European Language Resources Association, Magazzini del Cotone Conference Center, Genoa, Italy, May 24th – 26th, 2006.

Martínez-Rojas, M., Marín, N., & Vila, M. A. (2013). A preliminary approach to classify work descriptions in construction projects. *IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), 2013 Joint*. IEEE, Edmonton, Canada, June 24th – 28th, 2013. pp.1090-1095.

Matthews, P.H. (2002). *What can we conclude?* In R. M.W Dixon & Alexandra Y. Aikhenvald (eds.), Word: a cross-linguistic typology, Cambridge University Press, Cambridge, UK. pp. 266-281.

Morris, C. (1938). *Foundations of the theory of signs.* The University of Chicago Press, Chicago, IL, United States. <http://www.jstor.org/stable/4175798?seq=12#page_scan_tab_contents> (May 6th, 2016).

Newson, T.A., & Norbert J. D. (2011). Case methods in civil engineering teaching. *Canadian Journal of Civil Engineering.* 38 (9). pp. 1016-1030.

Palmer, D.D. (2000). *Tokenisation and sentence segmentation.* In: Dale R, Moisl H, Somers H, editors. Handbook of natural language processing. Marcel Dekker, New York, United States. pp. 11-35.

Poon, H., & Domingos, P. (2009). Unsupervised semantic parsing. *Proc., 2009 conference on empirical methods in natural language processing (EMNLP 2009)*. Singapore, August 6th – 7th, 2009.

Python v2.7.3 [Computer software]. Beaverton, OR, Python Software Foundation.

Rawlings, A., Paliokosta, P., Maisey, D., Johnson, J., Capstick, J., & Jones, R. (2014). A study to investigate the barriers to learning from serious case reviews and identify ways of overcoming these barriers. Kingston University Institute for Child Centred Interprofessional Practice, UK.

Richardson, L. (2015). Beautiful Soup. <https://www.crummy.com/software/BeautifulSoup/> (Jun 23rd, 2016).

Russell, M.A. (2011). *Mining the social web.* O'Reilly Media, Inc., Sebastopol, CA, United States. pp. 248-249.

Sparks, J.R., & Rapp, D.N. (2010). Discourse processing - examining our everyday language experiences. *Wiley Interdisciplinary Reviews: Cognitive Science*. 1 (3). pp. 371-381.

Salama, D.M., & El-Gohary, N.M. (2013). Semantic text classification for supporting automated compliance checking in construction. *J. Comput. Civ. Eng.* 10.1061/(ASCE)CP.1943-5487.0000301.

Santorini, B. (1990). Part-of-speech tagging guidelines for the Penn Tree-1192 bank project (3rd Revision, 2nd printing). <https://catalog.ldc.upenn 1193 .edu/docs/LDC99T42/tagguid1.pdf> (Jun 17th, 2016).

Schoenborn, J.M. (2012). A case study approach to identifying the constraints and barriers to design innovation for modular construction. M.Sc. Thesis. Virginia Polytechnic Institute and State University, Blacksburg, VA, United States.

Singh, A., & Sakamoto, I. (2001). Multiple Claims in Construction Law: Educational Case Study. *Journal of Professional Issues in Engineering Education and Practice*. 127 (3). pp. 122-129.

Shirazi, A. & Amir H.B. (2014). Design and assessment of a mobile augmented reality-based information delivery tool for construction and civil engineering curriculum. *Journal of Professional Issues in Engineering Education and Practice.* 141 (3). 04014012.

Thomas, H.R. & Ellis Jr, R.D. (2007). Interpreting construction contracts: fundamental principles for contractors, project managers, and contract administrators. American Society of Civil Engineers, Reston, VA, United States.

Thomas, H. R. (2009). Teaching Construction Contracts. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction.* 1 (4). pp. 165-168.

Toutanova, K., Klein, D., Manning, C., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. *Proc., HLT-NAACL 2003.* Edmonton, Canada, May 27th – June 1st, 2003. pp. 252-259.

Weekley, E. (1917). The romance of words. John Murray, London, UK. <https://babel.hathitrust.org/cgi/pt?id=umn.319510021804462;view=1up;seq=11> (May 6th, 2016).

Yin, R. (1984). *Case study research: design and methods.* SAGE Publications, Beverly Hills, CA.

Zhang, J. & El-Gohary, N. (2015a). Semantic NLP-based information extraction from construction regulatory documents for automated compliance checking. *J. Comput. Civ. Eng.* 10.1061/(ASCE)CP.1943-5487.0000346, 04015014.

Zhang, J. & El-Gohary, N. (2015b). Automated information transformation for automated regulatory compliance checking in construction. *J. Comput. Civ. Eng.* 10.1061/(ASCE)CP.1943-5487.0000427, B4015001.

Zhang, J. & El-Gohary, N. (2016). Extending building information models semiautomatically using semantic natural language processing techniques. *J. Comput. Civ. Eng.* 10.1061/(ASCE)CP.1943-5487.0000536, C4016004.

Zhou, P. & El-Gohary, N. (2015). Ontology-based multilabel text classification of construction regulatory documents. *J. Comput. Civ. Eng.* 10.1061/(ASCE)CP.1943-5487.0000530, 04015058.