Concept Relation Extraction from Construction Documents Using Natural Language Processing

Mohammed Al Qady[1] and Amr Kandil, M.ASCE[2]

Abstract: The objective of this research is to present an innovative technique for managing the knowledge contained in construction contract documents to facilitate quick access and efficient use of such knowledge for project management and contract administration tasks. Knowledge Management has become the focus of a lot of scientific research during the second half of the 20th century as researchers discovered the importance of the knowledge resource to business organizations. Despite early expectations of improved document management techniques, document management systems used in the construction industry have failed to deliver the anticipated performance. Recent research attempts to utilize analysis of the contents of documents to improve document categorization and retrieval functions. It is hypothesized that natural language processing can be effectively used to perform document text analysis. The proposed system, technique for concept relation identification using shallow parsing (CRISP), utilizes a shallow parser to extract semantic knowledge from construction contract documents which can be used to improve electronic document management functions such as document categorization and retrieval. When compared with human evaluators, CRISP achieved almost 80% of the average kappa score attained by the evaluators, and approximately 90% of their *F*-measure score.

DOI: 10.1061/(ASCE)CO.1943-7862.0000131

CE Database subject headings: Information management; Contract management; Information systems; Construction management; Computer aided operations.

Author keywords: Information management; Contract management; Information systems; Construction management; Computerization.

Introduction

Knowledge Management (KM) has become the focus of a lot of scientific research during the second half of the 20th century as researchers discovered the importance of the knowledge resource to business organizations. This importance is demonstrated throughout the various organizational levels by the emergence and use of terms such as "the knowledge society" and "the knowledge worker" (Drucker 1993) and the development of concepts such as "the learning organization" (Senge 1990). The way an organization manages knowledge involves how knowledge is created or extracted in the organization, preserved and communicated for effective utilization (Chinowsky and Molenaar 2005; Turk 2007; Walters et al. 2007).

To survive and succeed in gaining an advantage in the knowledge-intensive and highly competitive construction industry, effective use of large amounts of knowledge from various knowledge sources is essential. Realizing this fact, many construction firms started adopting and implementing some form of knowledge management system.

Edwards et al. (1996) estimated that approximately 80% of explicit construction knowledge is embedded in documents. A similar figure was reported by Tseng (2005). Construction projects produce vast amounts of documents. Turk et al. (1994) estimated that the construction of a single structure can generate about 10,000 documents. Moreover, numerous types of construction documents are generated and require effective management for successful contract administration: (1) contract documents; (2) correspondences; (3) minutes of meetings; (4) periodic progress reports; (5) quality and safety reports; (6) change order documents; (7) payment requisitions; (8) weather records; (9) material and equipment records; (10) employee time cards; (11) delay records; (12) records of overhead costs; and (13) claim documents (Rubin et al. 1999).

The importance of documents as a source of knowledge in construction projects, the large amount of documents produced by projects, and the diversity of such documents are factors that highlight the importance of document management for construction contract administration. Despite such importance and despite the expected increase in the use of advanced document management systems, document management practices in the construction industry have been described as inefficient and of limited reliability and cost-effectiveness (Chassiakos and Sakellaropoulos 2008; Lee et al. 2003) and inadequate and unable to provide the expected impact (Vidogah and Ndekugri 1998a).

The focus of this study is on the information retrieval (IR) aspect of document management systems. The concept relation identification using shallow parsing (CRISP) system was developed for the automatic/semiautomatic extraction of semantic information from the text of contract documents using a natural language processing (NLP) tool. It is anticipated that the semantic information contained in textual construction documents can be

[1]Research Assistant, Division of Construction Engineering and Management, School of Civil Engineering, Purdue Univ., West Lafayette, IN 47907. E-mail: malqady@purdue.edu

[2]Assistant Professor, Division of Construction Engineering and Management, School of Civil Engineering, Purdue Univ., West Lafayette, IN 47907 (corresponding author). E-mail: akandil@purdue.edu

Note. This manuscript was submitted on January 23, 2009; approved on July 18, 2009; published online on August 10, 2009. Discussion period open until August 1, 2010; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Construction Engineering and Management***, Vol. 136, No. 3, March 1, 2010. ©ASCE, ISSN 0733-9364/2010/3-294–302/$25.00.

used to improve document categorization, search and retrieval. In addition, recent research proposed the use of ontologies (El-Diraby and Kashif 2005; Lame 2004) for knowledge management. In this regard, the proposed tool can be used to assist in the knowledge acquisition stage of ontology building. Although the system was originally developed for document management applications in construction firms to facilitate contract administration processes, CRISP is not limited to a specific domain and can be used on any textual documents.

In the next sections, a brief review of contract administration, document management practices, and NLP is given, followed by a description of the semantic representations proposed in this study and the methodology employed to extract such knowledge from contract documents. The evaluation process of the developed system is then presented in addition to the results obtained from the evaluation. The final section gives a brief discussion on the results of the evaluation and demonstrates how the proposed system can be used to enhance document management activities.

Background

The contract administration process involves the management and organization of daily project tasks, including monitoring of project progress, project payment quantification, project change management, project time control and delay analysis, and project closeout processes. As part of contract administration duties, thorough documentation and record keeping procedures need to be implemented (Rubin et al. 1999). These documentation and record keeping procedures are viewed as an essential step to the prevention or defense of construction claims and disputes (Peña-Mora et al. 2003). The large amount of construction documents require effective document management systems that can cross reference important documents across topics of relevance and provide efficient access to these documents for multiple project stakeholders (Rubin et al. 1999). This need gave rise to research that attempted to develop document management systems in construction projects (Hajjar and AbouRizk 2000).

Traditional document management practices are based on the use of paper documents (Luiten et al. 1998; Stewart and Mohamed 2004). Several studies noted an expected increase in the use of electronic document management systems (EDMSs) resulting from a decline in the cost of hardware, an increase in the use of computers, and an increase in the computer proficiency of users (El-Tayeh and Gil 2007; Kangari 1995; Vidogah and Ndekugri 1998b). Although various studies outline the basic functions performed by an EDMS [e.g., Björk (2006); Vidogah and Ndekugri (1998b); Zipf (2000)], Turk et al. (1994) provided the following comprehensive list: (1) electronic archiving of documents; (2) creating, modifying, and printing documents; (3) getting or referencing external documents; (4) providing document confidentiality and security; (5) management of the relationship among documents; and (6) extracting documents or data from documents.

As explained in the previous section, several researchers have expressed their dissatisfaction with the EDMSs used for construction projects. Many EDMSs simply mimic the paper-based document management process (Zhu et al. 2007). In some, significant user input is required to define the relevant information for the system which then acts as a viewer for manually predefined information. In other words, the document processing effort is not transferred to the EDMS and remains a burden on the user. In addition, many EDMSs only act as digital archives with limited keyword-based search and retrieval capabilities (Fruchter et al. 2003).

Some recent studies attempted to utilize content analysis of documents to improve document categorizing and retrieval. Caldas et al. (2002) and Caldas and Soibelman (2003) used IR via text mining techniques to facilitate information management and permit knowledge discovery through automated categorization of various construction documents according to their associated project component. Fruchter et al. (2003) used text analysis to develop vector models of knowledge items which can then be retrieved according to their similarity with a user-defined search term. Meziane and Rezgui (2004) defined vector models for documents based on the documents' index terms and use term frequency and inverse document frequency to determine the similarity between two documents. In other words, index terms of the documents are used to represent the semantic knowledge in the documents.

In this study, NLP techniques are used to identify such semantic knowledge. The origins of language processing can be traced back to the 1940s (Jurafsky and Martin 2000). The artificial intelligence (AI) attribute of NLP was emphasized by the Turing Test (Turing 1950). In this test, a human interrogator tries to determine through questioning which of the other two concealed participants is a person and which is a machine. The late 1990s saw significant changes in the development of NLP techniques with the advent of probabilistic methods to refine the various algorithms (parsing, tagging, etc.), the emergence of the Web as a huge repository of text documents (which highlighted the importance of textual IR), and rapid advances in computer technology (Jurafsky and Martin 2000). Of particular importance to the present research is a group of NLP algorithms that perform a process called parsing. Parsing is a method for analyzing sentences and determining their grammatical structure (Allen 2003). In traditional parsing algorithms, the main task is to search through the space of all possible combinations of grammatical structures, to find the correct structure for a given sentence (Jurafsky and Martin 2000). This can cause the developed NLP applications to be less than efficient since sometimes such searches provide more information than needed for a particular application (Hammerton et al. 2002). The potential for inefficiency brought about another class of parsing algorithms, known as shallow or partial parsers. Shallow parsing algorithms have the task of recovering only limited information from sentences in natural language (Hammerton et al. 2002). This approach was found to be successful in reducing the ambiguity of sentences with multiple possible parses (Allen 2003). In addition, shallow parsers have been found to reduce search spaces, which increase their robustness in applications such as question-answering and speech-to-speech translation (Hammerton et al. 2002).

Semantic Representation of Contractual Knowledge

The objective of the developed CRISP technique is to extract contractual knowledge and create a semantic representation of this knowledge. Therefore, before delving into the details of the CRISP methodology, this section will illustrate the form of the semantic representation used to represent contractual knowledge. Semantic information expressed in documents can be used to improve document management functions used for contract administration tasks by construction firms. To achieve this, semantic information must be extracted from the documents and structured into a formal representation which can be applied to document

J. Constr. Eng. Manage., 2010, 136(3): 294-302

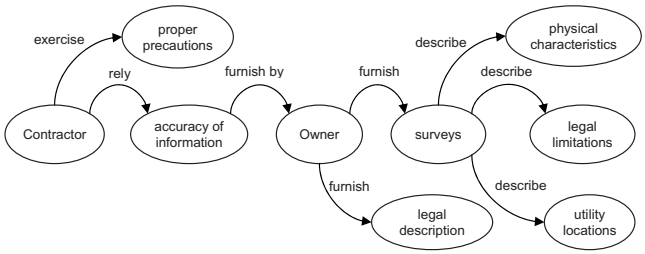

Fig. 1. Semantic modeling of a sample contract clause

management related tasks. In this study, the object-oriented approach described by Gomez-Pérez (1998) is used, in which concepts are identified by nouns and relations identified by verbs. Accordingly, a "concept set" is a set made up of the following three components:

- Active concept: the concept doing the action or the subject of the relation;
- Relation: the action or the verb of the relation; and
- Passive concept: the concept that is being acted upon or the object of the relation.

For example, suppose the following sentence: the contract documents form the contract for construction. The active concept is the "contract documents," the relation is "form" and the passive concept is the "contract for construction." In semantic notation, the sentence can be expressed as follows: form (contract documents, contract for construction), where the active concept contract documents and the passive concept contract for construction are related by the relation form. Based on this semantic representation, the following standard contract clause can be modeled, as shown in Fig. 1.

> The owner shall furnish surveys describing physical characteristics, legal limitations, and utility locations for the site of the project and a legal description of the site. The contractor shall be entitled to rely on the accuracy of information furnished by the owner but shall exercise proper precautions relating to the safe performance of the work.

Concept Relation Identification Using Shallow Parsing—Technique

The CRISP technique utilizes an NLP tool to automatically/semiautomatically identify and extract possible concept sets from construction contract documents. The framework of CRISP is illustrated in Fig. 2. A computer program was developed in *C++* to manipulate the various components of the system for the purpose of extracting the concept sets.

Input File Preparation

In order for the program to successfully analyze a document presented to it, the document must first be prepared in a standard format that can be read by the program. For evaluation purposes, two file preparation strategies were developed:

- Basic preparation: basic preparation defines the structure of the document to the program. Contract documents are usually divided into sections, subsections, clauses, subclauses, etc. For IR purposes and to facilitate evaluation of the system, it is important to tag each concept set extracted by the program with the section number it was extracted from. Section numbers in the input file are bound by angle brackets to identify to the program that the sentences following the section number are part of that specific section and, consequently, all concept sets extracted from these sentences are tagged with the appropriate section number. In addition, all sections of the document are ended with the dummy sentence "clause_end" to identify to the program the boundaries of each section. Section titles, if available, are not included in the input file; and
- Advanced preparation: advanced preparation, which also includes the basic preparation steps, is used to resolve the issues created by enumerations and lists in the text. Sentences containing enumerations and lists are elaborated to facilitate accurate parsing of the sentences. Basically the same process is used to resolve both enumerations and list; the sentence is broken down into separate sentences, each containing a component of the enumeration or the list. List numbering, if present, is also removed. This process is a systematic manual process that does not look at the grammatical structure of the resulting sentences. In other words, the resulting sentences after the resolution of enumerations or lists may contain grammatical errors. No effort is made to correct such errors; CRISP relies on the ability of the NLP tool to correctly parse ungrammatical sentences.

Shallow Parsing

Unlike previous research in building construction ontologies in which conceptualization was a manual process [e.g., El-Diraby and Kashif (2005)], concept sets in this study will be extracted either automatically or semiautomatically using NLP techniques, namely, a shallow parser. It was realized from the onset of the project that the accuracy of the system will be highly sensitive and highly dependent on the accuracy of the NLP tool that is used. Accordingly, the choice of the shallow parser was an im-

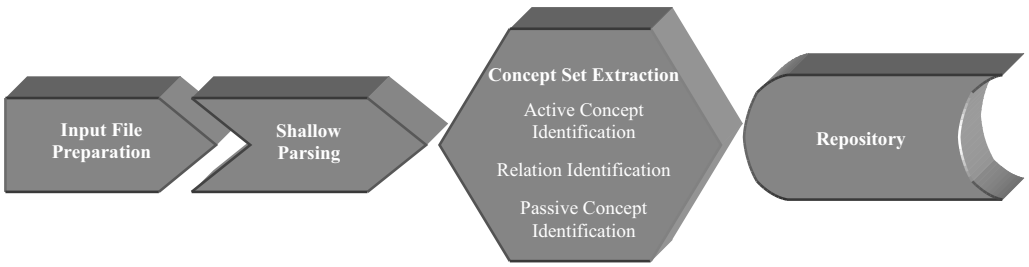

Fig. 2. CRISP framework

J. Constr. Eng. Manage., 2010, 136(3): 294-302

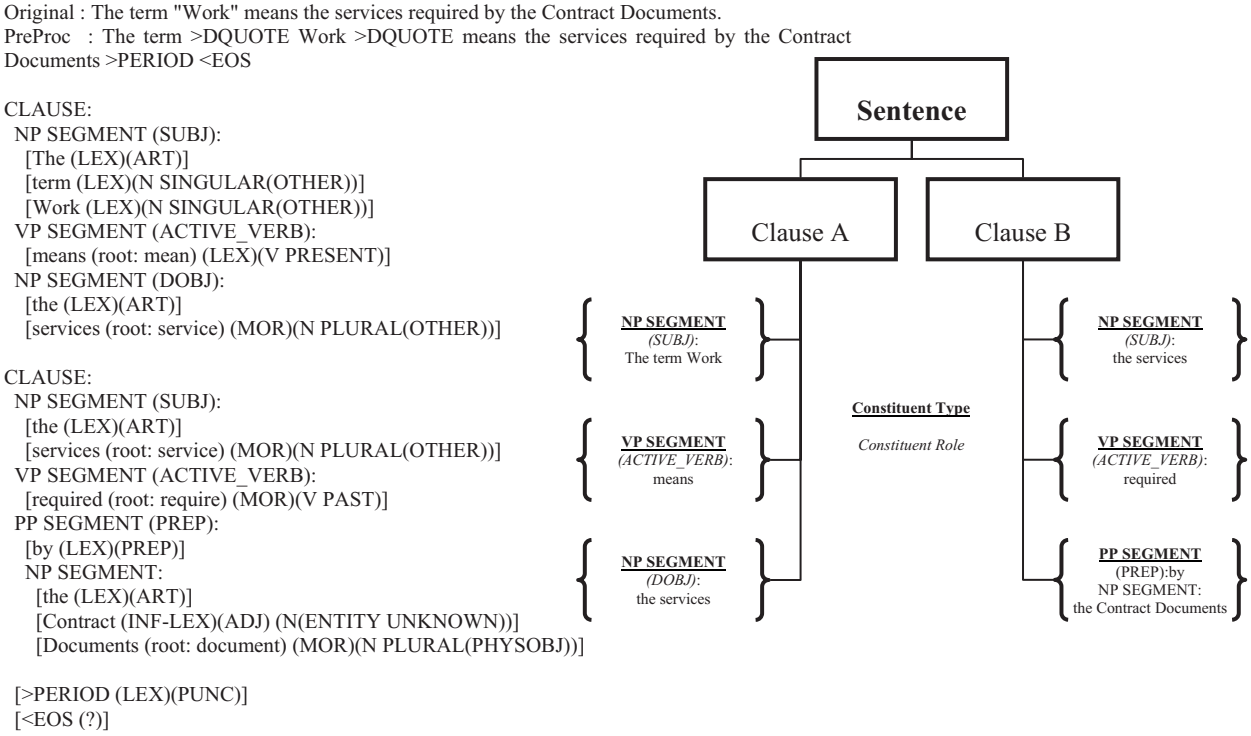

Fig. 3. Sample Sundance parse

portant decision; the shallow parser sentence understanding and concept extraction (Sundance) was used. Sundance is a natural language parser developed by the School of Computing at the University of Utah (Riloff and Phillips 2004). Sundance is a shallow parser because sentences processed by Sundance are parsed into a two-level structure. The first level is the clause level in which the sentence is broken down into either a single clause or multiple clauses depending on the sentence structure. The second level is the constituent level in which words in each clause are grouped into either noun phrases (NPs), verb phrases (VPs), prepositional phrases (PPs), or adjective phrases. Fig. 3 gives a sample parse produced by Sundance and the conceptual form of the sentence demonstrating its final two-level structure. As shown in the figure, the constituents at the second level are labeled by their types (NP segment, VP segment, PP segment) and their roles, if any, are identified (SUBJ, DOBJ, and ACTIVE_VERB). By having this simple structure, a computer program can easily access constituents at either of the two levels by looping through the sentence; a single loop will access the clause level, two nested loops will access the constituent level. For example, Clause B can be accessed in the second pass of a single loop while the PP "by the contract documents" can be accessed using a double nested loop, in the second pass of the first loop and the third pass of the second loop.

Sundance was chosen for its ability to perform syntactic segmentation and assign syntactic roles to NPs. Also, previous research in the field of textual case-based reasoning used Sundance to index legal cases and develop case representations that can be compared to identify similarities between legal cases and even predict outcomes of the cases (Brüninghaus and Ashley 2005; Brüninghaus and Ashley 2001). Sundance has been described as a robust state-of-the-art parser by the researchers and was considered a useful resource for their work. The choice of Sundance for the implementation of CRISP was also based on a review of a number of available NLP toolkits. For example, the Stanford Parser was used for a number of sample parses. Due to the advantages afforded by shallow parsing, as previously discussed, it was decided that deep parsing was not necessary and that a shallow parser was sufficient. However, the writers would like to mention that, as part of their future research, other shallow parsers may be experimented with as well due to the rapid developments that take place in this field of NLP.

Concept Set Extraction

The three components of a concept set are identified as follows:

- Active concept: the segments under each clause are checked for a constituent with role "SUBJ." If the subject segment is identified, it is extracted and considered the active concept component of the concept set. Constituents following the subject segment are checked to identify PPs that follow the subject. PPs following the subject are considered to qualify the active concept and are therefore extracted along with the subject segment;
- Passive concept: the segments under each clause are checked for a constituent with role "DOBJ." If the object segment is identified, it is extracted and considered the passive concept component of the concept set. Constituents following the object segment are checked to identify PPs that follow the object. PPs following the object are considered to qualify the passive concept and are therefore extracted along with the object segment. In some cases, no direct object is identified in the clause. This especially occurs when the VP is followed by a PP. In such cases, where all clause segments are checked and none have a constituent role of DOBJ, the constituent type for the segment directly after the VP is checked. If it is a PP then it is extracted and assumed to be the passive concept component of the concept set; and
- Relation: Sundance divides sentences into clauses based on the existence of multiple VPs and, consequently, each clause will

J. Constr. Eng. Manage., 2010, 136(3): 294-302

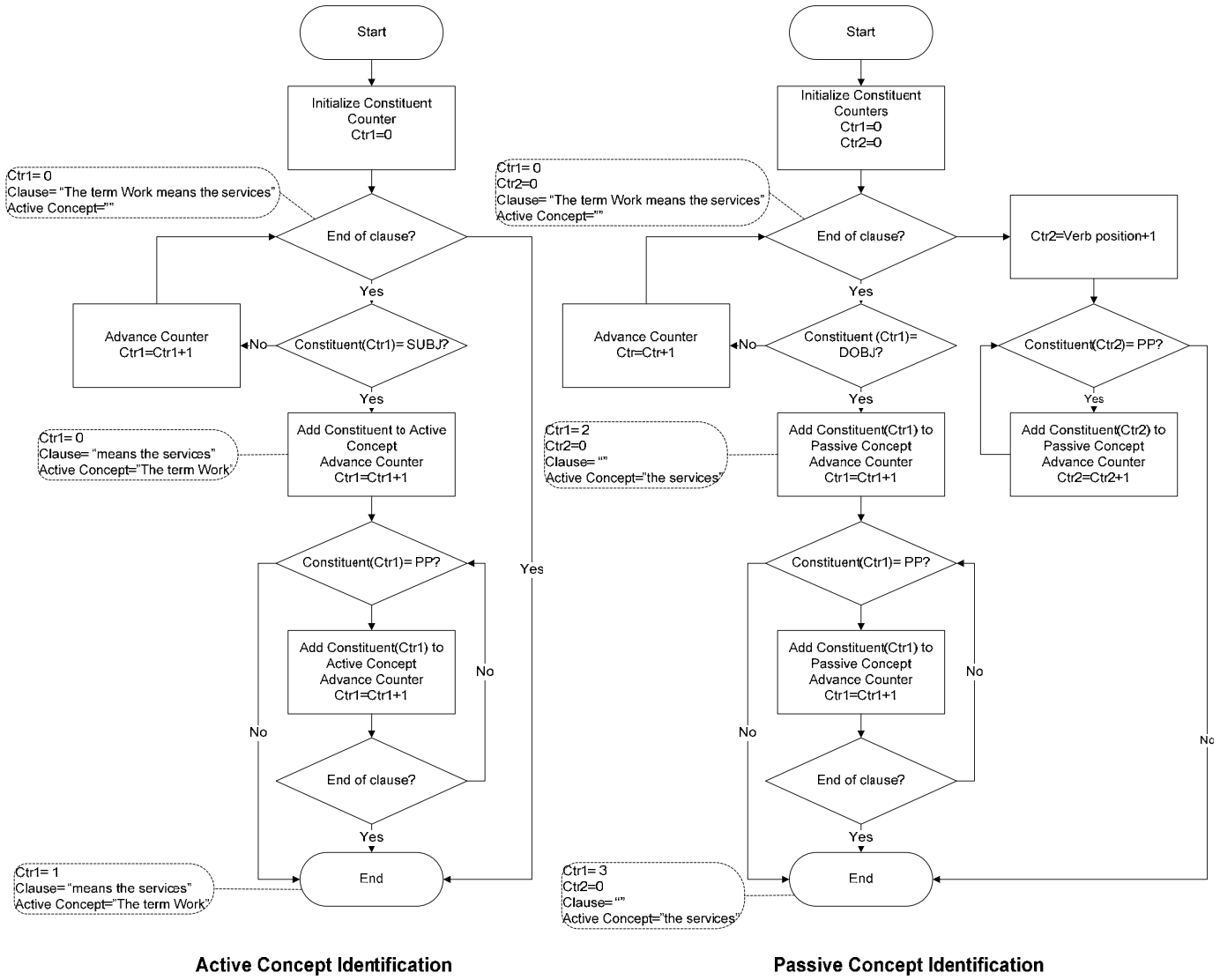

Fig. 4. Active concept and passive concept extraction algorithms

have only one VP. The segment under each clause with constituent type "VP segment" is extracted and considered the relation component of the concept set.

Fig. 4 presents a flowchart for the processes used in identifying both the active concept and the passive concept.

Repository

All concept sets extracted from the document are stored in the repository. The repository is consequently the subject of the system's evaluation and will be the main component for using the system for document management-related tasks.

Evaluation and Results

The first step in the evaluation procedure is to choose an input document to be analyzed by CRISP. Construction contract documents usually include a variety of textual documents expressed in natural language (e.g., the agreement, the general conditions of the contract, the specifications, etc.). It was decided to conduct the test on a standard form of contract, namely, the American Institute of Architects (AIA) Document A201-1997, general conditions of the contract for construction [American Institute of Architects, Inc. (AIA) 1997], for the following reasons:

- Standard forms of contract are widely accepted by all project participants and widely used in the construction industry. Accordingly, a meaningful analysis of standard forms can be beneficial to all parties using the standard form;
- Because they undergo numerous revisions, standard forms are practically error free; and
- AIA documents have evolved over almost 115 years through numerous editions to become benchmark documents expressing the contractual relationships between construction parties (American Institute of Architects 2008). AIA Document A201-1997 is the 15th edition of the document, with the first edition dating back to 1888.

Characteristics of the Input File

A201 is made up of 14 articles, each article dealing with an important topic in construction projects such as general provisions, changes in the work, payments and completion, etc. Articles are made up of sections. Sections may be divided into subsections,

J. Constr. Eng. Manage., 2010, 136(3): 294-302

and subsections may be divided into subsubsections. All in all, A201 is made up of 264 provisions, a total of 19,679 words, with an average of 75 words per provision.

An important feature found in standard forms of contract in general and in A201 in specific is the presence of enumerations and lists in the provisions' text. The following sentence is a sample enumeration from the text of A201:

> the owner shall furnish surveys describing physical characteristics, legal limitations, and utility locations for the site of the project and a legal description of the site.

The following sentence is a sample list from the text of A201:

> a modification is (1) a written amendment to the contract signed by both parties; (2) a change order; (3) a construction change directive; or (4) a written order for a minor change in the work issued by the architect.

Preliminary experiments with Sundance using sentences containing enumerations and lists showed that the accuracy of the parser in dividing a sentence into clauses decreases and, accordingly, syntactic role assignment is affected. In addition, enumerations and lists produce multiple active concepts, relations, and/or passive concepts in a single sentence clause. Therefore, even if the clause segmentation was correct, Sundance will identify only one NP with role SUBJ and only one NP with role DOBJ per clause, ignoring the multiplicity. The extent of the impact of parser accuracy is clearly demonstrated in the following section of the paper in the difference between the results obtained for Output 1 versus Output 2. Since the performance of the whole system depends largely on the accurate assignment of syntactic roles and due to the abundant use of lists and enumerations in the input document, this text feature had a critical effect on the performance of CRISP. Based on this finding, it was decided to prepare a modified input file and evaluate CRISP on both an original input file (prepared using the basic preparation steps) and a modified input file (prepared according to the advanced preparation steps), as described in the previous section. Although, advanced preparation could be a tedious process, especially when sentences contain multiple enumerations or combined enumerations and lists that result in numerous possible combinations, the purpose of preparing a modified input file was to evaluate CRISP's performance independent of the accuracy of the shallow parser and compare how the system will perform with and without the parser's handicap.

Evaluation Process

The concept sets extracted by CRISP from the two input files were compared to the concept sets extracted by human evaluators to determine precision, recall, and agreement with human evaluators. Because it is unpractical to have human evaluators evaluate all 264 provisions of A201, an evaluation set of provisions had to be prepared. To avoid biased selection, a random number generator was used to select 15 of the 264 provisions. To get comprehensive and accurate results, it was important that the evaluation exercise should not be a heavy burden on the human evaluators. Accordingly, out of the 15 provisions, six were selected according to the following criteria:

- To avoid any bias, the full provision is either selected or it is excluded from the evaluation set; partial provisions were not used;
- The average words per provision of the whole document is 75, accordingly, the evaluation set must have an equivalent average words per provision;
- The selection of provisions must not be from a certain part of A201. The chosen provisions must be distributed over the whole body of the document; and
- The provisions must include the common features found in standard forms of contract such as enumerations and lists.

Both original and modified input files were prepared for the evaluation set and processed by the program. The output of CRISP was prepared in a spreadsheet format to facilitate comparison with human output from the evaluation session.

A 2-h evaluation session was arranged for seven human evaluators. All evaluators were graduate students in civil engineering with practical experience in the construction industry. All of them had previously completed at least one course in construction contracts, the majority completing two or more. The amount of experience each evaluator had ranged from 1 to 6 years in construction management and contract administration positions that range from entry to intermediate level. A short 10-min presentation was given to outline the objective of the exercise, explain what concept sets are and demonstrate the extraction of active concepts, relations, and passive concepts from simple sentences. The evaluation exercise was then presented to the evaluators and the evaluators were instructed to try to be as comprehensive as possible in identifying the concept sets. The results of the human evaluators were compiled and prepared in a spreadsheet format.

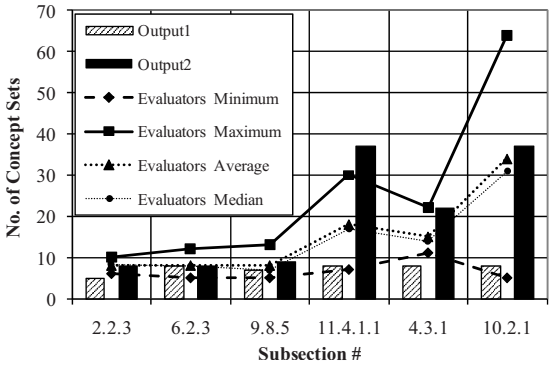

Fig. 5. Number of extracted concept sets

Preliminary Results

In terms of the number of concept sets extracted by each evaluator and by CRISP for both input files, Fig. 5 shows that the number of concept sets extracted by the system from the original input file (Output 1) was generally on the lower boundary of the numbers extracted by the human evaluators while the number of concept sets extracted from the modified input file (Output 2) was generally around the average number extracted by human evaluators. These results also highlight the subjectivity of the exercise: despite the fact that the same provisions were evaluated by all the human evaluators, the results show significant variations in the number of concept sets extracted by the evaluators in most of the provisions. Although these numbers show that CRISP extracted concept sets within the expected limits, the number of concept sets does not reflect the performance of the system. For an accurate evaluation of performance that takes into account the subjec-

J. Constr. Eng. Manage., 2010, 136(3): 294-302

Table 1. Kappa Scores

Evaluator	A	B	C	D	E	F	G	Average evaluators (%)
A	—	28%	40%	39%	35%	29%	43%	36
B	28%	—	47%	29%	28%	30%	48%	35
C	40%	47%	—	48%	38%	48%	76%	50
D	39%	29%	48%	—	37%	37%	47%	40
E	35%	28%	38%	37%	—	37%	38%	36
F	29%	30%	48%	37%	37%	—	57%	40
G	43%	48%	76%	47%	38%	57%	—	52
Average evaluators	36%	35%	50%	40%	36%	40%	52%	41
Output 1	10%	20%	23%	19%	19%	17%	21%	19
Output 2	34%	31%	40%	29%	30%	24%	38%	32

tive nature of the investigation, kappa scores and precision and recall values were calculated for the results of the human evaluators and both outputs of the system.

Kappa

Kappa measures the pairwise agreement between two human evaluators or a human evaluator and the system's output after adjusting for chance agreement (Jurafsky and Martin 2000). Kappa is calculated according to Eq. (1), where $P(A)$ is the probability of agreement and $P(E)$ is the probability of chance agreement

$$k = \frac{P(A) - P(E)}{1 - P(E)} \tag{1}$$

In the above equation, $P(A)$=proportion of concept sets agreed on by both evaluators to the total number of unique concept sets identified by both evaluators and $P(E)$=probability that a concept set agreed on by both evaluators was chosen at random and is therefore equal to the square of the inverse of the number of unique concept sets identified by both evaluators. The result of each human evaluator was compared in a pairwise comparison with the results of the other human evaluators and the results of CRISP. All in all, there were 35 comparisons; 21 human-human comparisons and 14 human-computer comparisons.

Kappa was calculated between the concept sets extracted for each subsection by each pair of evaluators. An average kappa over the six subsections is calculated representing the average agreement between the two evaluators over the complete evaluation set. Finally, an average kappa is calculated over the 21 human-human comparisons which is compared to the average kappa of the seven human-computer comparisons. Table 1 presents the detailed results of this evaluation technique. The results emphasized the subjectivity of the investigated issue; on average, human evaluators agreed only 41% of the time. Output 1 of the system achieved slightly less than 50% of the performance achieved by the human evaluators, with an average agreement of 19% with humans. This number increased to 32% with Output 2, which corresponds to achieving almost 80% of the performance of the human evaluators.

Fig. 6 shows the average agreement achieved by each individual evaluator with the other human evaluators and the system's agreement with each individual evaluator. The figure demonstrates how both computer outputs followed the general agreement trend of the human evaluators. For example, evaluators C and G achieved the highest average kappa scores among all human evaluators. Likewise, both computer outputs achieved their highest kappa scores in comparisons with evaluators C and G.

F Measure: Precision and Recall

Precision, or accuracy, measures the percentage of correct concept sets from the total number of concept sets extracted by the evaluator/system while recall, or coverage, measures the percentage of correct concept sets extracted by the evaluator/system from the total number of correct concept sets. To measure precision and recall, concept sets extracted by the human evaluator and by the system must be identified as either "correct" or "incorrect." The gold standard represents the truth, in this case the concept sets that are deemed to be correct. However, because of the subjectivity of the results of the study, the gold standard is not readily apparent and must therefore be determined from the results of the human evaluation. The arithmetic total of the concept set extracted by the human evaluators was 640. The number of repetitions of each concept set extracted by the evaluators was identified. Repetition values ranged from a maximum of seven (for a concept set identified by all the evaluators) to a minimum of one (for a concept set identified by only one evaluator). Considering repetition, the actual total number of concept sets was determined at 314. The next step was to determine a threshold of repetitions according to which a concept set is considered correct. A low threshold increases precision and reduces recall while a high threshold decreases precision and increases recall. In determining the threshold, the kappa scores achieved by the evaluators were considered. High agreement between the evaluators would justify the use of a high threshold. However, since evaluator agreement proved to be relatively low, a threshold that is slightly toward the low end was used. A concept set is considered correct if at least three evaluators agreed on the concept set. Concept sets

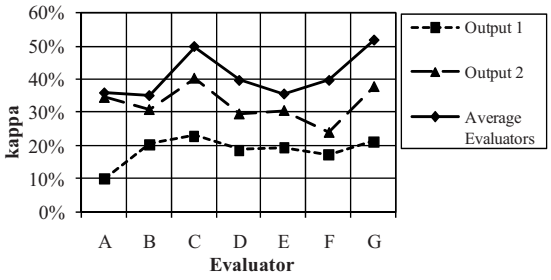

Fig. 6. Average kappa scores

Table 2. Precision, Recall, and *F*-Measure Scores

Evaluator	Precision (%)	Recall (%)	*F* measure (%)
A	83	61	71
B	56	86	68
C	90	93	92
D	90	59	71
E	73	63	68
F	71	71	71
G	84	98	90
Average evaluators	78	76	76
Output 1	48	38	42
Output 2	70	67	68

that satisfy these criteria were gathered to develop the gold standard. Of the 314 concept sets, 71 concept sets made it to the gold standard.

Precision and recall were calculated by comparing the results of the human evaluators and computer outputs with the gold standard for each subsection in the evaluation set. Precision and recall values are averaged over the six subsections in the evaluation set to determine the average precision and recall of a human evaluator or a computer output. *F* measure is used to combine precision and recall into one measure (Jurafsky and Martin 2000). *F* measure is calculated according to Eq. (2), where *P* is the precision value and *R* is the recall value. β is a parameter that is used to assign relative weights to the precision and recall values. At this stage, precision and recall are considered of equal importance; accordingly, the value for β is one

$$F \text{ measure} = \frac{(\beta^2 + 1)PR}{\beta^2 P + R} \tag{2}$$

Table 2 lists the precision, recall, and *F*-measure scores for all seven human evaluators and both computer outputs. Compared with the gold standard, the human evaluators achieved an *F*-measure score of 76% while Output 1 of the system achieved an *F*-measure score of 42%. Quite similar to the kappa results, Output 1 achieved slightly more than 50% of the performance achieved by the human evaluators. The *F*-measure score of the system increased to 68% with Output 2, corresponding to approximately 90% of the performance achieved by the human evaluators.

Summary and Conclusions

Human interpretation of written text varies from one person to the next depending on many factors such as knowledge, experience, priorities, and effort. This observation holds even when the evaluated text is well structured and with minimum ambiguity such as standard forms of contract. Accordingly, the evaluation of a computer system that tries to resemble humans in their linguistic abilities is not straightforward since there is no clear baseline to judge the performance of the system against.

Two methods were used for evaluating the performance of CRISP, a computer system that tries to emulate humans in their ability to identify related concepts from text expressed in natural language. In the first method, the degree of agreement of the human evaluators over the investigated task was measured to identify a threshold for what is expected of human performance. The degree of agreement of the computer system with the human evaluators was then compared to this threshold to determine the level of human performance the system was able to achieve. In the second method, a baseline or gold standard was developed from the results of the human evaluation, and both the human evaluators and the computer system were compared to this baseline to determine to what degree the performance of the computer system resembled the performance of the human evaluators. The degree of agreement of the human evaluators as determined from the first evaluation method was used for guidance in developing the gold standard.

Results obtained by the system can be considered encouraging. The standard agreement level between human evaluators was 41%. The system in its best output achieved an agreement level of 32%, approximately 80% of the average human performance. Moreover, the system displayed agreement trends that were equivalent to the trends observed from the human evaluations. In terms of precision and recall, the average *F* measure of all human evaluators based on the established gold standard was 76%. The best *F* measure achieved by the system was 68%, approximately 90% of the average human performance. The fact that both evaluation methods produced relatively comparable results testifies to the validity of both evaluation methods.

As expected, the performance of the system was highly dependent on the performance of the shallow parser used. A review of the incorrect concepts extracted by the system revealed two main parsing errors:

- Errors resulting from complex features in the input files, namely, enumerations and lists; and
- Errors resulting from incorrect syntactic segmentation, namely, the incorrect identification of nouns as verbs and the consequent formation of incorrect VPs.

Advanced input file preparation was introduced in an attempt to reduce the effect of the first parsing error. Advanced preparation may be a manual and tedious process. However, the logic behind the idea was to try to evaluate the system's performance independent of Sundance's specific inaccuracies and determine whether or not a little bit of effort in input file preparation will be rewarded with improved performance. Indeed, a significant improvement in performance was observed:

- 75% increase in kappa scores; and
- 62% increase in *F*-measure scores.

Correction of the second parsing error required significant recoding of Sundance to modify the heuristics employed by the successive phrase segmenters. Accordingly, it was decided to ignore this error (especially since its effects on the results were not as drastic as the first error's effects) thereby absorbing its effects into the performance of the overall system.

Although the purpose for developing the system was to enhance document management techniques in the construction industry, CRISP can be used on any textual documents and is not limited to a specific domain. Various document management applications can benefit from the use of a system like CRISP. As previously discussed, more research is focusing on text analysis instead of the traditional keyword search methods. Subject to further investigations, it is anticipated that document categorization and retrieval based on the analysis of semantic relationship in the text will yield better results in comparison with traditional keyword-based techniques. CRISP can also be used as a component of an IR system to enable querying in natural language. As discussed above, recent research has been focusing on the use of ontologies to facilitate KM. Ontologies not only provide a common foundation for knowledge exchange but can also be used to enable natural language communication with computers. CRISP can be used to assist in the difficult and time-consuming process

J. Constr. Eng. Manage., 2010, 136(3): 294-302

of extracting concept and concept relation from texts, thus assisting in the knowledge-acquisition stage of ontology building. Finally, CRISP can be used to develop visualizations of the important concept and concept relations in text documents. In contract documents, this can be used, for example, to illustrate the roles of the major parties of the contract which can be helpful for educational and training purposes.

Acknowledgments

The research team would like to thank Ellen Riloff and Siddharth Patwardhan from the University of Utah for providing us with a copy of Sundance. Special thanks to Nick Pendar from Iowa State University for his valuable assistance. This study is supported by the National Science Foundation (Award No. NSF-CMMI-0700363). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the writers and do not necessarily reflect the views of the National Science Foundation.

References

Allen, J. (2003). *Natural language understanding*, 2nd Ed., Benjamin-Cummings, Redwood City, Calif.

American Institute of Architects. (2008). "History of contract documents." ⟨http://www.aia.org/docs_history⟩ (March 2008).

American Institute of Architects, Inc. (AIA). (1997). "General conditions of the contract for construction." *AIA A201*, Washington, D.C.

Björk, B. C. (2006). "Electronic document management in temporary project organisations: Construction industry experiences." *Online Inf. Rev.*, 30(6), 644–655.

Brüninghaus, S., and Ashley, K. (2005). "Reasoning with textual cases." *Case-based reasoning research and development*, Springer, Berlin, 137–151.

Brüninghaus, S., and Ashley, K. D. (2001). "The role of information extraction for textual CBR." *Proc., 4th Int. Conf. on Case-Based Reasoning*, Springer, Berlin, 74–89.

Caldas, C. H., and Soibelman, L. (2003). "Automating hierarchical document classification for construction management information systems." *Autom. Constr.*, 12(4), 395–406.

Caldas, C. H., Soibelman, L., and Han, J. (2002). "Automated classification of construction project documents." *J. Comput. Civ. Eng.*, 16(4), 234–243.

Chassiakos, A. P., and Sakellaropoulos, S. P. (2008). "A web-based system for managing construction information." *Adv. Eng. Software*, 39(11), 865–876.

Chinowsky, P., and Molenaar, K. (2005). "Learning organizations in construction." *Proc., Construction Research Congress 2005: Broadening Perspectives*, ASCE, Reston, Va., 839–848.

Drucker, P. R. (1993). *Post-capitalist society*, Butterworth-Heinemann, Stoneham, Mass.

Edwards, D. J., Shaw, T., and Holt, G. D. (1996). "Electronic document management systems and the management of UK construction projects." *Build. Res. Inf.*, 24(5), 287–292.

El-Diraby, T. E., and Kashif, K. F. (2005). "Distributed ontology architecture for knowledge management in highway construction." *J. Constr. Eng. Manage.*, 131(5), 591–603.

El-Tayeh, A., and Gil, N. (2007). "Using digital socialization to support geographically dispersed AEC project teams." *J. Constr. Eng. Manage.*, 133(6), 462–473.

Fruchter, R., Demian, P., Yin, Z., and Luth, G. (2003). "Turning A/E/C knowledge into working knowledge." *Proc., 4th Joint Int. Symp. on Information Technology in Civil Engineering*, ASCE, Reston, Va., 143–155.

Gomez-Pérez, A. (1998). "Knowledge sharing and re-use." *Handbook of applied expert systems*, J. Liebowitz, ed., CRC, Boca Raton, Fla.

Hajjar, D., and AbouRizk, S. (2000). "Integrating document management with project and company data." *J. Comput. Civ. Eng.*, 14(1), 70–77.

Hammerton, J., Osborne, M., Armstrong, S., and Daelemans, W. (2002). "Introduction to special issue on machine learning approaches to shallow parsing." *J. Mach. Learn. Res.*, 2, 551–558.

Jurafsky, D., and Martin, J. H. (2000). *Speech and language processing*, Prentice-Hall, Upper Saddle River, N.J.

Kangari, R. (1995). "Construction documentation in arbitration." *Constr. Engrg. and Mgmt.*, 121(2), 201–208.

Lame, G. (2004). "Using NLP techniques to identify legal ontology components: Concepts and relations." *Artif. Intell. Law*, 12(4), 379–396.

Lee, H.-S., An, S.-J., Son, B.-S., Jang, M.-H., and Choi, Y.-K. (2003). "Web-based electronic data interchange model to improve the collaboration of participants in construction projects." *Proc., Construction Research Congress 2003: Winds of Change*, ASCE, Reston, Va., 871–879.

Luiten, G. T., Tolman, F. P., and Fischer, M. A. (1998). "Project-modelling in AEC to integrate design and construction." *Comput. Ind.*, 35(1), 13–29.

Meziane, F., and Rezgui, Y. (2004). "A document management methodology based on similarity contents." *Inf. Sci. (N.Y.)*, 158(1–4), 15–36.

Peña-Mora, F., Sosa, C. E., and McCone, S. D. (2003). *Introduction to construction dispute resolution*, 1st Ed., Prentice-Hall, Upper Saddle River, N.J.

Riloff, E., and Phillips, W. (2004). "An introduction to the Sundance and AutoSlog system." ⟨http://www.cs.utah.edu/~riloff/pdfs/official-sundance-tr.pdf⟩ (August 2007).

Rubin, R., Fairweather, V., and Guy, S. (1999). *Construction claims prevention and resolution*, 3rd Ed., Wiley, New York.

Senge, P. M. (1990). *The fifth discipline: The age and practice of the learning organization*, Century Business, London.

Stewart, R., and Mohamed, S. (2004). "Evaluating web-based project information management in construction: Capturing the long-term value creation process." *Autom. Constr.*, 13(4), 469–479.

Tseng, F. S. C. (2005). "Design of a multi-dimensional query expression for document warehouses." *Inf. Sci. (N.Y.)*, 174(1–2), 55–79.

Turing, A. M. (1950). "Computing machinery and intelligence." *Mind*, LIX, 433–460.

Turk, Ž. (2007). "Construction informatics in European research: Topics and agendas." *J. Comput. Civ. Eng.*, 21(3), 211–219.

Turk, Z., Bjork, B. C., Johansson, K., and Sevensson, K. (1994). "Document management systems as an essential step towards CIC." *Proc., CIB W78 Workshop on Computer Integrated Construction*, International Council for Building Research Studies and Documentation, Helsinki.

Vidogah, W., and Ndekugri, I. (1998a). "Improving the management of claims on construction contracts: Consultant's perspective." *Constr. Manage. Econom.*, 16(3), 363–372.

Vidogah, W., and Ndekugri, I. (1998b). "Review of the role of information technology in construction claims management." *Comput. Ind.*, 35(1), 77–85.

Walters, R., Jaselskis, E. J., and Kurtenbach, J. M. (2007). "Classification of knowledge within the electrical contracting industry: A case study." *Leadership Manage. Eng.*, 7(1), 11–17.

Zhu, Y., Mao, W., and Ahmad, I. (2007). "Capturing implicit structures in unstructured content of construction documents." *J. Comput. Civ. Eng.*, 21(3), 220–227.

Zipf, P. J. (2000). "Technology-enhanced project management." *J. Manage. Eng.*, 16(1), 34–39.