

## Rule-based NLP Methodology for Semantic Interpretation of Impact Factors for Construction Claim Cases

Jia Niu<sup>1</sup> and Raja R.A. Issa<sup>2</sup>

<sup>1</sup>Rinker School of Construction Management, University of Florida, PO Box 115703, Gainesville, FL, USA32611-5703; PH(352)328-7483; email: [jniu@ufl.edu](mailto:jniu@ufl.edu)

<sup>2</sup>Rinker School of Construction Management, University of Florida, PO Box 115703, Gainesville, FL, USA32611-5703; PH(352)273-1152; email: [raymond-issa@ufl.edu](mailto:raymond-issa@ufl.edu)

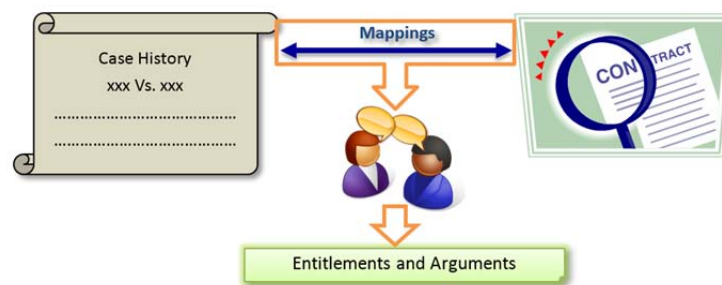
### ABSTRACT

Construction claim analysis largely depends on determining the existence of impact factors. Extensive research has been conducted on the relationships between impact factors and outcomes of construction litigation from various perspectives and using various methodologies. However, only a few of them have taken into consideration the automated semantic interpretation of impact factors contained in a construction litigation case. In this paper, based on previous pilot studies on the domain ontologies of construction contractual semantics, a rule-based NLP (Natural Language Processing) methodology for semantically interpreting impact factors for construction claim cases is proposed. Based on the available NLP techniques and domain ontologies, this methodology utilizes a rule-based mechanism to achieve the mapping from textual elements to ontology entities. In this way, the support of domain ontology is provided to enhance the performance of the impact factor interpretation process in the text. Also, it was found that several software packages can work together to satisfy the demand for the implementation of this methodology. Further, to test the validity of this methodology, several case studies focusing on DSC (Differing Site Conditions) claims are conducted by adopting cases from legal databases as data. The significance of this research is that it provides a more automated functionality to the traditional approach of claim outcome prediction by adding one more semantic factor-interpreting layer to it, and by also exploring the application of ontology-based NLP in the domain of construction claim analysis via text processing.

### INTRODUCTION

In construction contractual management, contractual claims analysis is an important aspect that can significantly impact a project's profit or loss under certain circumstances. In a claim case, the domain experts' legal analysis is the core part in generating justifiable claim entitlements and arguments. Generally speaking, the classic work flow of legal analysis in construction claim cases can be distilled as the

model shown in Figure 1. In this figure, the domain experts have to read through the history of a claim case and determine the facts happened on a project. At the same time, they should also be aware of the legal interpretation of the articles and clauses of the contract and the related domain knowledge and insights. Subsequently, the task that they should perform is to find the mappings between the facts that happened in the project history and the corresponding clauses in the contract. Based on the mappings found, the experts can identify the entitlements for the client and then develop arguments to generate compensation claims. Thus, the mappings linking facts and domain knowledge are the key elements in this model, and effectively and efficiently finding those mappings is essential in claim practice.



**Figure 1. Work flow model for the legal analysis of construction claim cases**

The current claims practice paradigm for finding the mappings adopts impact factors associated with certain types of claims to modulate and simplify this process. Diekmann and Nelson (1985), Kululanga et al. (2001) and Arditi and Thanat (2008) have conducted research based on this paradigm. Most of their work involved the analysis and collection of impact factors through different methodologies, and studying the relationships between the varied combination of impact factors and the litigation outcomes. Almost all of these studies use an assumption that the existence of impact factors in the text of a claim case's history should be manually interpreted and determined, i.e. a human has to read through and interpret the text and then decide whether each of the predefined impact factors exist in a given case.

This paper proposes bypassing this manual process by using a rule-based NLP (Natural Language Processing) methodology supported by domain ontology, for semantically interpreting impact factors in the text of a construction claim case history file.

## BACKGROUND

As mentioned above, based on the adoption of impact factors in construction claims and disputes, many significant progress has been achieved in claims management. One of the areas of research that has shown such progress is litigation

outcome prediction. Around this goal, initially impact factors were extracted, quantified and measured from precedent litigation cases by using a variety of methods (Mahfouz and Kandil 2009 and 2012). The relationships between the factors and outcomes in precedent cases were also found by machine learning methods (Mahfouz and Kandil 2010 and 2012). Subsequently, a multi-agent system was developed to generate legal arguments (Kandil 2010). This system used the impact factors as input data with a formal logic algorithm to simulate the process of legal discourse and obtain a resolution to a construction dispute.

While these studies were successful to a large extent, there still exists considerable room to improve the level of automation. Although the claim outcome can be predicated based on the impact factors as input data, the interpretation and determination of the existence of those factors are thrown back to the end users, analogous to the criticism to those experts system developed in 1980s to 1990s (Bubbers 1991). It is very common that legal concepts have a characteristic of inexactness. For example, in Differing Site Conditions claims, “materially different”, “act as a reasonable prudent contractor”, “reasonable reliance” and “reasonably unforeseeable” are the crucial concepts for judging the impact factors’ existence. As a matter of fact, it is not an easy job for a person lacking enough domain knowledge to correctly understand the semantics of those concepts and make a correct decision. Usually to interpret the concepts for judging an impact factor’s existence one has to follow a rigorous set of logical rules regarding the varied combinations of conditions and scenarios in the real-world. Hence, the purpose of this research is to automatically interpret the semantics for the impact factors in the text of claims cases.

To solve this problem, an ontology-based methodology for construction claim domain knowledge representation and semantic interpretation was proposed in a line of pilot studies (Niu and Issa 2012). Briefly, the roadmap of this solution lies in representing the domain knowledge about facts and conditions for judging the existence of impact factors by using ontology, and using that ontology as the intelligent support with the text processor for identifying language patterns to determine whether any impact factors appear in a claim case.

Following this roadmap, considerable progress has been made in previous pilot studies using domain ontology development (Niu and Issa 2013a, 2013b, 2013c). Also, for the text processor in this approach, NLP techniques can be considered as a solution. Basically, NLP allows for the application of a streamline of basic text processing steps (Tokenization, Sentence Splitting, POS (Part-Of-Speech) tagging, Gazetteer compiling etc.) to implement the tasks of entity and relation extraction (Cunningham et al. 2011), which are the basis for this proposed methodology. But, the question of how to integrate the NLP-enabled text processor to work with the developed domain ontology is next question to be studied, and is discussed in the next section.

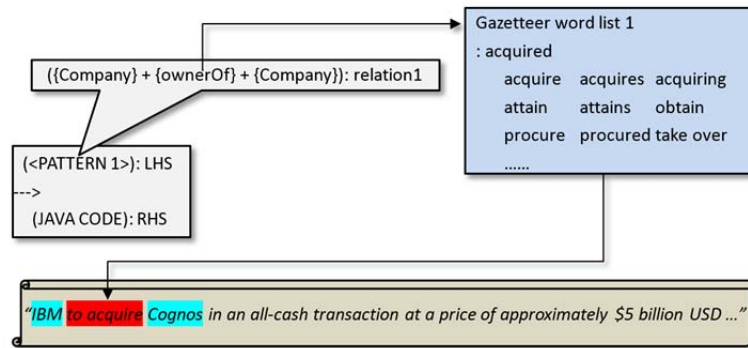
## METHODOLOGY

To fulfill the objective of capturing the mappings between the patterns of facts in case history text and the impact factor, in the proposed methodology, rules are used as a model for explicitly holding a mapping. Within a rule, patterns stay on the left hand side and an impact factor on the right. Particularly for the left hand side of a rule, the NLP technique goes straight to the text, and processes it literally and syntactically through performing entity extraction and relation extraction. By NLP alone some mappings could be captured at a shallow level, but to achieve the full potential of this methodology, an ontology should be available to provide semantic reference and intelligent support. Thus, the key issue in this regard is how to identify another kind of mapping from extracted textual elements to ontology entities. In the following, the detailed mechanisms of these three important parts of this proposed methodology are explained by using examples.

**Entity Extraction.** Assume that a fact that happened in the case history of a construction claim can be broken down and simplified somehow into a triple structure like “subject-predicate-object” (W3C 2004). Within this structure, both subject and object can be treated as entities and these entities follow certain syntactic rules to represent certain meanings. The syntactic rules can be described by the NLP’s tasks of gazetteer types and POS tagging, and further the entities can be captured by those rules and marked out by annotations. For instances, a token with an NNP (Proper Noun) POS tag followed by an annotation for Company suffix like “Ltd” or “LLC” would be annotated as an entity of type Company or Organization; type “title” followed by “firstname” and “lastname” annotations would be annotated entirely as an entity of type Person, like “Mr John Smith”.

**Relation Extraction.** Following that triple mode, a predicate can be deemed as a relation between a subject and an object. The NLP is then capable of capturing a relation by using gazetteer types and pattern matching rules. Figure 2 shows the mechanism for capturing a relationship in a sentence. The vertical scroll shape represents the text from a claim case history file. To capture the relation of “ownerOf” shown in the callout on the left hand side of a rule, it has to satisfy two conditions. One is literally, a match of instance within the gazetteer (word list) for “acquired” should be found in the text, i.e. “to acquire” which is highlighted in red in the text. Another condition is a syntactic one that the match should come with two instances of gazetteer type of “Company” both in the front and in the end respectively, i.e. “IBM” and “Cognos” highlighted in green. Also, this syntactic condition can be defined as “relation1” appearing on the left hand side of a rule, as shown in the callout shape in the Figure 2.

**Rule-Based Mapping from Annotations to Ontology Classes.** Although entities and relation extraction by NLP alone can capture all three elements in a triple structure for a fact, its shortage of semantic information largely constrains the interpretation capability. It is necessary to connect the extracted entities or relations to the ontology classes in order to acquire domain knowledge like the attributes of an entity, and the relationships between two entities.



**Figure 2. Illustration for the mechanism of relation extraction**

To accomplish these connections, rules have to be explicitly defined to describe the mappings. For example, in Figure 3, "Drawing A502, Exterior Doors and Windows Details" has been captured and annotated as in the "Drawings" class after entity extraction was performed. Then, to establish this link, in left hand side of a rule, a pattern matching grammar is used to match this annotation "Drawings" to the corresponding ontology class "Drawings" as a subclass of "ContractDocument". Subsequently, due to the inheritance in the taxonomy, the subclass "Drawings" inherits the attribute of "being legally binding" from "ContractDocument" (American Institute of Architects 2007a, 2007b). As a result, this semantic information for the entity of "Drawing A502, Exterior Doors and Windows Details" is established.

**Implementation Tools.** Last but not the least, through a study on common NLP development and ontology editor software, the author discovered that certain software packages can be used together as effective tools to satisfy the demand for the implementation of this methodology (Horridge 2011; Thakker et al. 2009). Those tools and purposes are listed in Table 1.

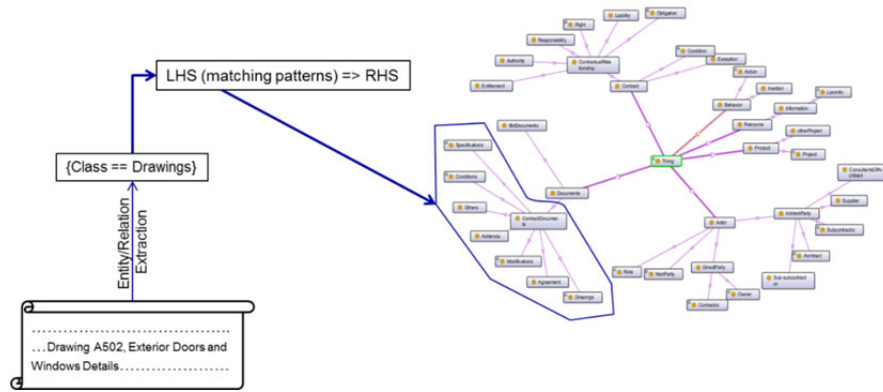


Figure 3. Mapping from textual elements to ontology entities

CASE STUDY

**Objective.** A case study was designed to apply this proposed methodology, and to validate the objective of this case study, which is to identify the factors impacting the outcome of certain claim types. To prevent involving redundant variables, it was determined to focus on one of the most common claim types and to keep the number of impact factors in a manageable range. Considering previous studies and cases involving claim impact factors, Differencing Site Conditions (DSC) claims and the classic 6 factors for Type I and 3 factors for Type II DSC claim (International Technology Corporation v. Donald C. Winter, Secretary of the Navy 2008; Renda Marine, Inc. v. United States 2007; Weeks Dredging & Contracting, Inc. v. United States 1987) were selected as the objective of this case study.

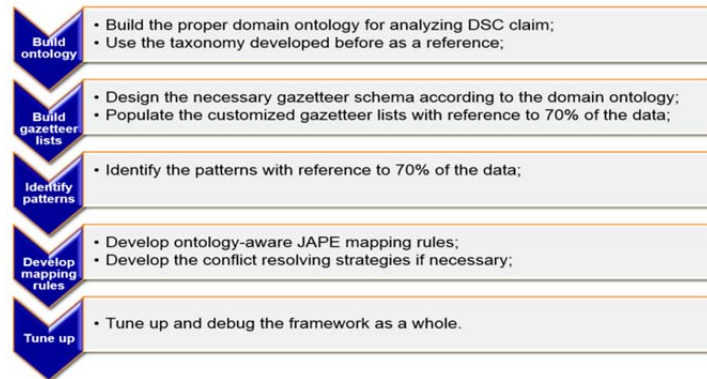
Table 1. Implementation Tools

Tools	Purposes
Protégé 4.3	Ontology engineering environment
GATE Developer 7.1	NLP development environment
ANNIE	Plug-ins for NLP tasks, like tokenization, sentence splitting, POS tagging, etc.
JAPE	Rules for pattern matching from annotations to ontology
Java	Object-oriented Programming Language for doing certain manipulations

**Data Collection and Preparation.** The data needed is the precedent DSC claim cases in construction projects. Basically, domain ontology, gazetteers and patterns should be learned from these cases. To collect data, a professional legal database, such as the *LexisNexis Academic* database, were used. In *LexisNexis Academic*, the results returned from searching with “DSC” as the keyword can be used as raw data. However, some preparatory work should be conducted on this raw data to make the data ready for use, in order to discard the irrelevant cases and also the text should be reformatted and simplified. Seventy percent of the data will then be used for

developing the system, and the rest (30%) for the system validation process.

**Execution.** To utilize the proposed methodology in this case study, the essential modules should be completed in a certain sequence. The main activities have been arranged in an execution plan as illustrated in Figure 4.



**Figure 4. Execution Plan of the Case Study.**

## SIGNIFICANCE AND FUTURE WORK

The significance of this research manifests itself twofold. From the academic perspective, this study explores the application of ontology-based NLP in the domain of construction claim legal analysis via text processing, and the domain ontology to be developed can serve as a reusable structure for organizing concepts of the domain knowledge about construction contractual and claim. From a practical point of view, one more layer of semantic factor-interpreting was added to the traditional way of construction claim outcome prediction. This additional layer could also work with the claim outcome predicting system to improve the level of automation.

For the future work, validation is an important issue. It is necessary to set up a “Gold Standard” as the benchmark for measuring the performance of the case study analyses. Also, future work should extend the claim types processed beyond DSC claims to include other more common claim types of claims.

## REFERENCES

- American Institute of Architects. (2007a). *AIA Document A201 - 2007 Commentary*. Washington, D.C.
- American Institute of Architects. (2007b). *General Conditions of the Contract for Construction, AIA A201*. Washington, D.C.
- Arditi, D. and Pattanakitchamroon, T. (2008). Analysis Methods in Time-Based Claims. *Journal of Construction Engineering and Management*, ASCE, 134(4), 242-252.
- Bubbers, Geoffrey R. (1991). *Computer assisted construction claim analysis*. M.Sc.

- Thesis, University of New Brunswick, NB, Canada.
- Cunningham et al. (2011). *Text Processing with GATE (Version 6)*, University of Sheffield, UK, Department of Computer Science.
- HorrIDGE, Matthew. (2011). *A Practical Guide To Building OWL Ontologies Using Protege 4 and CO-ODE Tools*, The University of Manchester, 1.3 ed..
- International Technology Corporation v. Donald C. Winter, Secretary of the Navy, 523 F.3d 1341; 2008 U.S. App. 2008.
- Kandil, Islan H. El-Adaway and Amr A. (2010). Multiagent System for Construction Dispute Resolution (MAS-COR). *Journal of Construction Engineering and Management*, ASCE, 136, No.3, 303-315.
- Mahfouz, T., and Kandil, A. (2009). Factors Affecting Litigation Prediction of Differing Site Conditions (DSC) Construction Cases: A Discrete Binary Choice Model. *2009 Construction Research Congress*, Seattle, Washington.
- Mahfouz, T., and Kandil, A. (2012). Litigation Outcome Prediction of Differing Site Condition Disputes through Machine Learning Models. *Journal of Computing in Civil Engineering*, ASCE, 26(3), 298-308.
- Mahfouz, T., and Kandil, A. (2010). Construction Legal Decision Support Using Support Vector Machine. *2010 Construction Research Congress*, Banf, Alberta, Canada.
- Niu, Jia, and Issa, Raja R. A. (2012). Framework for Production of Ontology-based Construction Claim Documents. *ASCE International Conference on Computing in Civil Engineering*, ASCE, Clearwater Beach, Florida.
- Niu, Jia, and Issa, Raja R. A. (2013a). Concept Classification for Construction Contractual Semantics: A case study on the AIA A201 Document. *30th CIB W78 International Conference*, CIB, Peking, China.
- Niu, Jia, and Issa, Raja R. A. (2013b). Conceptualizing Methodology for Building an Ontology for Construction Claim Knowledge. *ASCE International Conference on Computing in Civil Engineering*, ASCE, Los Angeles, California.
- Niu, Jia, and Issa, Raja R. A. (2013c). Taxonomy for the Domain Knowledge of Construction Contractual Relationships between Owners and Contractors. *EG-ICE 20th International Workshop: Intelligent Computing in Engineering 2013*, EG-ICE, Vienna, Austria.
- Renda Marine, Inc. v. United States, 509 F.3d 1372; 2007 U.S. App. 2007.
- Thakker, D., Osman, T., and Lakin, P. (2009). *GATE JAPE Grammar Tutorial (1.0 Ed.)*.
- W3C. (2004). *Resource Description Framework (RDF): Concepts and Abstract Syntax*. 10 February 2004. <http://www.w3.org/TR/rdf-concepts/>
- Weeks Dredging & Contracting, Inc. v. United States, 13 Cl. Ct. 193; 1987 U.S. Cl. Ct. 1987.