Morphological Localization Solution in e-Learning

Martin Jovanović
Faculty of Electronic Engineering Niš
Aleksandra Medvedeva 14
18000 Niš, Serbia
+381 18 529 523
martin.jovanovic@elfak.ni.ac.rs

Dejan Todosijević
Faculty of Electronic Engineering Niš
Aleksandra Medvedeva 14
18000 Niš, Serbia
+381 18 529 443
todosijevic.dejan@gmail.com

ABSTRACT
This paper represents a follow-up to the first author's BCI 2013 paper on the topic of localization in an e-learning framework and as such it discusses the implementation of the solution proposed therein. The motivation for these papers is the problem of morphology in Western Balkans (or any other morphology-dependent) languages which imposes an additional layer of complexity onto e-learning systems that provide any form of natural language feedback to users. One such system has been developed at Computer Science Department of the University of Niš, Faculty of Electronic Engineering. This system provides elementary feedback in form of a relation between two arbitrarily chosen notions in the learning material. In this contest the returned statement, in the form of subject, predicate and object, often requires a case of the object (noun) that is different from the original case in the text. In languages that rely on morphology a different case means a different form of the word. Without the proper case the returned statement is still understandable; however, its improper grammar may be distractive to the learner. This obstacle has been overcome in a relatively simple fashion by additional semantics. This paper discusses the exact ways.

Categories and Subject Descriptors
D.3.1 [**Computer Uses in Education**]: Computer-Assisted Instruction (CAI) – *Semantic Web, RDF and drag-and-drop user interface in the context of learning* The ACM Computing Classification Scheme: http://www.acm.org/class/1998/

General Terms
Algorithms, Design, Languages.

Keywords
e-learning, CAI, semantic web, instructional design, RDF, drag and drop, interface, DSi.

1. INTRODUCTION
Though the primary motivation for this paper and the underlying research is linguistic in nature, it is distraction-free e-learning that is in focus of both. The primary goal of e-learning is facilitation, acceleration and enrichment of learning by the means of computer technology. Design of e-learning systems usually assumes, other than technology, the spatial and often temporal distance between the student and the teacher [2]. This approach opens up a vast array of new possibilities, to the extent of an entire new learning paradigm – the one in which the driving force for learning shifts from the teacher (the *push* approach) to the student (the *pull* approach), as well as the quantity (the corporate-beloved "just enough" principle) and the timing ("just in time" or "on demand" learning). [3] In order to provide required contents in a required timeframe, large scale e-learning systems often store the learning material in relatively small units ("learning objects") which are combined and sequenced dynamically – often with the aid of other technologies, like ontologies [4] or other Semantic Web tools. [5] Semantic web turned out to be a perfect supplement to e-learning [6] as it provides not only means for structuring, aggregation and organization of learning material, but also a better personalization of e-learning experience by providing means for student modeling. All this, as well as the high degree of scalability built in into the core of Semantic web paradigm, drove the development of the DSi e-learning framework, the central topic of this paper, to some of the Semantic web technologies. Though this framework is in early stage of development and relatively simple in structure, the choice of the Semantic web approach guarantees its scalability.

In: A. Bădică, M. Colhon (eds.): Proceedings of the 2015 Balkan Conference on Informatics: Advances in ICT

2. THE DSI PROBLEM
As the Drag-and-Drop Semantic Interface e-learning framework (DSi) has been thoroughly discussed at the BCI conference, [1] in this paper only the brief review of its features will be given. The framework has been prototyped [7] with the idea of high scalability (by using Semantic Web tools) and graph-based knowledge representation (inspired by the idea of concept maps). [8]

2.1 DSi Framework Review
The DSi framework consists of two layers: the textual layer (learning material in plain or rich text) and the semantic layer (the graph of notions from the text interconnected by relations between them). The latter is given in the form of RDF/XML. [9]

On page load, the text in not sent to the client immediately. It is first parsed for any words that exist in the semantic (RDF) document. Once matches are found, all those words are encapsulated in span tags which are provided with drag and drop capability. This way all the words that exist in the RDF document can be dragged and dropped onto in the text displayed to the learner. This is the first important stage in the framework lifecycle (Figure 1).

When the user drags one word and drops it onto another (the word can only be dropped onto another drag gable word, Figure 2), the

dragged and the dropped-onto word are sent to the server, where the RDF document is searched for any RDF statement that includes both of these words as statement subject and object, respectively.

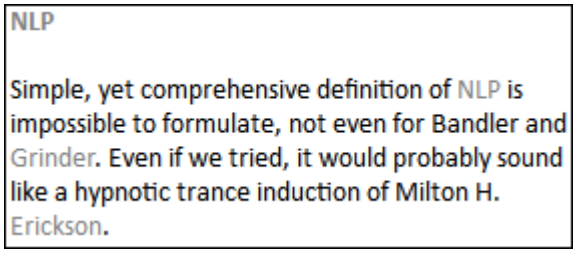

Figure 1. Lecture text with draggable words.

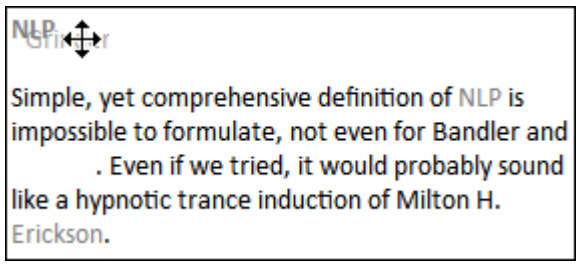

Figure 2. Drag and drop operation.

If one or more such statements exist, the server will return statement predicates in all of them. These predicates are relations between the two words user has chosen (Figure 3).

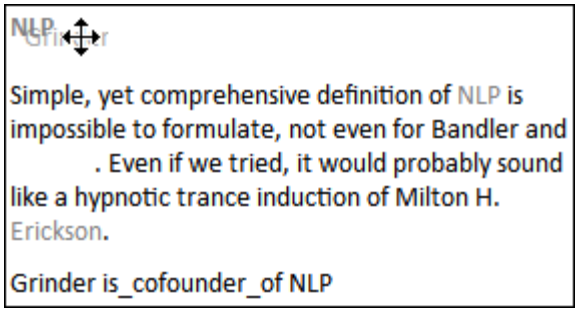

Figure 3. Resulting statement (relation).

The framework is, like mentioned before, in an early stage of development, thus there is a lot of space for improvements. For example, relations are given in a free, human language form, and as such do not carry any formal semantics – prone to automated reasoning or classification. However, the primary intent of this version is testing on human subjects and gaining knowledge on practical usability and initial shortcomings. One of these was the topic of [1], and another was encountered while the first one was overcome and tested.

2.2 Morphology Issue

The first usability issue was the question of morphological transformations of nouns in morphology-dependent languages, such as most of the Western Balkans languages. Though development of the framework in being carried in Serbian language environment, initial versions were implemented with the English learning text (as shown in Figures 1 and on). First time Serbian language was applied, the morphology issue emerged: the object word in a natural language sentence, almost always, has a different lexical form in the subject-predicate-object (SPO) statement than in the text. This is due to cases – in Serbian language there are 7. In order to have a lexical match between the word in the text and the word in the returned SPO statement, the word in the text must (happen to) be in the same case as in the SPO statement (in practice most frequently the 4th and 7th case). In a free-style natural language text this may and may not happen (the probability is, roughly, 1/7). The examples from [1] may illustrate this more clearly. In Serbian, nouns come in 7 cases (the word "stolica" – "chair" - is taken as an example):

1. nominative ("who?") - *stolic**a***,
2. genitive ("of ") - *stolic**e***,
3. dative ("Give to…") – *stolic**i***... etc.

Similar situation exists in Greek language, only there are 4 cases; technically, the issue is identical (*anthropos*, Greek for man):

1. *o anthrop**os***,
2. *tou anthrop**ou***... etc.

The word "stolica" (chair) can be found in any of 7 cases throughout the text. However, in a relation such as "laptop is on the chair" ("laptop je na stolic**i**") requires the word "stolica" to be in the 7th case ("stolic**i**"); there is no guarantee that this situation will happen in the natural text and there is now way to force it. Therefore, when a SPO statement is displayed to the user, it will most probably be grammatically incorrect.

3. THE SOLUTION

This issue has been solved by deepening the semantic layer of the document by one level. In order to describe this, a brief overview of the semantic layer will be given.

3.1 State of Affairs on the Semantic Side

Each textual lesson in DSi framework is accompanied by an RDF document. This document contains the mapping between certain notions from the text (technically – relations between certain words). For example, one RDF statement will define relations that the word "Erickson" has with all the words it's related to. This is best shown in the form of a graph (Figure 4).

In the RDF/XML syntax, this statement will look similarly to the Figure 5, with possible slight variations in code.

Relations are stated in a simple form – as an element of an RDF statement (or part of it). In order to define the required case for the object word in the statement, additional information must be provided in the RDF: *each predicate's required object-word case*. In order to achieve this, compound RDF statements had to be broken down into singular ones – so that each singular RDF statement (carrying only one relation) be enriched by this additional parameter – required object-word case.

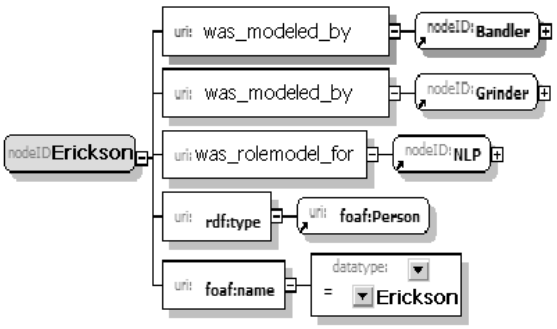

Figure 4. Relations in the RDF document.

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:eg="http://example.org/foovocab#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
<eg:Person rdf:nodeID="Erickson">
<eg:name>Erickson</foaf:name>
<eg:was_a_rolemodel_for rdf:nodeID="NLP"/>
<eg:was_modeled_by rdf:nodeID="Bandler"/>
<eg:was_modeled_by rdf:nodeID="Grinder"/>
</eg:Person>
</rdf:RDF>
```

Figure 5. RDF/XML syntax of a statement.

An example of the singular RDF statement, in the syntax of the latest revision of the framework, could look like in Figure 6. To emphasize the morphological aspect, an inverse relation has been given (instead of "Erickson was modeled by Bandler", the relation "Bandler modeled Erickson" was given as it provides a better example for morphology).

```
<rdf:Description rdf:about="Bandler">
  <modeled>Erickson</modeled>
</rdf:Description>
```

Figure 6. A singular relation.

In the initial framework version, when the word Bandler would be dropped onto the word Erickson, the system would return the statement "Bandler *modeled* Erickson". In Serbian, this would be:

Bandler *je_modelovao* Erickson.

This sentence requires the object word (Erickson) to be in the 4th case. However, in the text, this might not be the case. So, if the object-word is taken directly from the text, it will show in a wrong case in the statement returned from the system.

The initial solution proposed in [1] suggested that a new piece of information be added to each singular RDF statement: required case. The case was to be given by its name, though number or any ID would do. However, during the practical implementation somewhat different solution was proved to be more adequate.

3.2 Solution Implementation

Instead of adding required case into the singular RDF statement, the process which handles the drop operation took the different path:

1. When the word is dropped onto another, the two words (subject and predicate) are sent to the server.
2. Both words are turned into nominative (1st) case. This is done by searching all words in all cases; once the match is found, on any case, the nominative case of that word is returned.
3. Based on both words in 1st case the predicate is sought for (the same way as in the previous DSi version). [1]
4. Once the predicate is obtained, the adequate case for the object-word is determined by the case requirement statement (Figure 7).
5. Once the case is determined, the appropriate case of the object word is obtained from the additional RDF document which contains only cases for the words in the primary RDF and all object-word candidates from the text (one typical RDF statement from this document is given in Figure 8).

```
<rdf:Description rdf:about="modeled">
   <subject> Bandler </subject>
   <object> Erickson </object>
   <Erickson> genitive </Erickson>
</rdf:Description>
```

Figure 7. Case requirement statement.

```
<rdf:Description rdf:about="Erickson">
   <nominative>Erickson</nominative>
   <genitive>Ericksona</genitive>
   <dative>Ericksonu</dative>
   <accusative>Ericksona</accusative>
   <vocative>Ericksone</vocative>
   <instrumental>Ericksonom</instrumental>
   <locative>Ericksonu</locative>
</rdf:Description>
```

Figure 8. A case RDF document statement.

This way the statement returned to the user will always have a proper grammar. However, during the testing, another issue emerged.

4. ADDITIONAL PROBLEM

In the initial version of the framework the learning text was in English. In transition to Serbian, the case problem in the returned statements was spotted and the solution was implemented. During further testing, a new issue emerged: in Serbian texts the framework was unable to make all the words found in the RDF document draggable.

The sequence of the execution was the following:

1. split the text into separate words by traversing the DOM tree and focusing on terminal nodes;
2. search the RDF document for each word from the text;
3. in case of a match, make the word draggable and highlight it (with color).

Not all the words that existed in the RDF document were highlighted; even several instances of the same word were highlighted and some were not. This occurred due to the lexical comparison with the words in the RDF – while all the words in RDF statements are in the 1st case (nominative). Only nouns in the

1st case throughout the text were positively compared to the words in the RDF and thus marked as draggable. Words in any other case (with the exception of some specific words) we false negatives – the lexical comparison was negative due to the morphological transformations in cases.

4.1 Solution

The solution required a change in the execution sequence and comparison. The new RDF document, containing cases for object-candidate words, needed to be expanded to encompass both subject and object candidates.

On page load, any word from the text is compared against all cases of both subject and object candidate words; that part the execution sequence is altered (the system addresses the case RDF document instead of the relations RDF document). In this implementation all instances of all subject-candidate and object-candidate words were recognized, highlighted and spanned as drag-droppable.

This aspect was completely neglected when the case problem was first, tackled and it wasn't until the first practical testing of the solution proposed in [1] that the attention was drawn to it.

5. CONCLUSION

This paper discussed the solution to the localization problem in an e-learning system aimed at morphology-rich languages, such as Western Balkans region languages. The e-learning framework in question is labeled DSi (Drag and Drop Semantic Interface) [10] and provides means to retrieve relations between any two words in the text on the go, as the learning progresses, with no need for reviewing definitions at the beginning of the course. The framework is operational for learning materials in English or any other language with morphologically stable nouns. However, in languages that rely on morphology to convey the meaning (in which nouns morphologically alter depending on the context), the framework returns grammatically improper responses. The solution was proposed in [1] and implemented with some differences to the proposed way. Testing the solution revealed another issue of the exact same nature which caused the problems on an unexpected location – in the initial application loading phase, limiting the framework functionality (a more serious limitation than the improper grammar of the responses). This issue led to some framework implementation changes, generalizing the case RDF document application to both application load phase and user interaction (drag and drop action) handling.

Completing the application in this respect hasn't, however, resolved all the potential morphology-induced issues. Though fully operational for the current purposes, [11] the application might not operate properly with other morphological transformations, such as possession. This aspect is relatively easy to address in English (where possessions include "*'s*" suffix which can be tackled with a regular expression). However, possessions in Serbian are built morphologically (and as such can serve as nouns) and are also prone to case changes. Example would be "*Martin's*" which translates to "*Martinov*" for male, "*Martinova*" for female and "*Martinovo*" for neutral possession, all of which come in 7 forms (cases) and can, in certain situations, assume the role of nouns (thus be candidates for dragging and dropping). This forms an entirely new class of words to be addressed in further research.

Other possible research directions have already been stated in [1] and include enriching the relations with properties (transitive, reflexive or symmetrical), bringing relations themselves into relations etc.

6. REFERENCES

[1] Jovanović, M. - " Localization in E-learning Semantics (DSi Model Approach)", 6th Balkan Conference on Informatics BCI 13, Thessaloniki, Greece, September 19-21, 2013, Proceedings of the 6th Balkan Conference in Informatics, pp. 164-170, published by ACM New York, NY, USA ©2013, ISBN: 978-1-4503-1851-8

[2] Ally, M. – Foundations of Educational Theory for Online Learning, u Anderson, T. (editor) - The Theory and Practice of Online Learning, Athabasca University Press, 2008, ISBN 1897425082, pp. 15-44.

[3] Horton, W. – Leading E-Learning: Here is how You Can Chart your Course, Champion Implementation, Ensure Success, American Society for Training and Development, 2001, ISBN 1562862987.

[4] Fayed G., Sameh D., Ahmad H., Jihad A., Samir E. i Hosam E. - E-Learning Model Based On Semantic Web Technology, International Journal of Computing & Information Sciences, Vol. 4, No. 2, 2006, pp. 63–71.

[5] Semantic Web, W3C Standards Web page, accessible online at: http://www.w3.org/standards/semanticweb/ (URL last visited on March 31st, 2015).

[6] Pahl, C., Holohan, E.: Applications of Semantic Web Technology to Support Learning Content Development, Interdisciplinary Journal of E-Learning and Learning Objects, vol. 5, 2009.

[7] Jovanović, M. - "Arhitektura konceptualno-orijentisanog alata za podršku učenju", Infoteh Conference, Jahorina, Republic od Srpska, 28-30. 03. 2007, Infoteh-Jahorina Proceedings of Papers, Vol. 6, Ref. E-IV-4, pp. 477-479, published by Faculty of Electronic Eastern Sarajevo, R. Srpska, 2007, ISBN 99938-624-2-8.

[8] Novak, J. D. & A. J. Cañas, The Theory Underlying Concept Maps and How to Construct Them, Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition, 2008.

[9] RDF (Resource Description Framework) - W3C Specification available at: http://www.w3.org/RDF/ (URL last accessed on 1st of April, 2015.

[10] Jovanović, M. - "Semantička nadgradnja nastavnog materijala u sistemima za elektronsko učenje" ("Semantic Upgrade of Learning Material in E-Learning Systems"), M.Sc. thesis, University of Niš, Serbia, Faculty of Electronic Engineering, 2009.

[11] The DSi showcase application can be found at the following URL: http://www.martin.rs/DSi1.5_TestModule_v4/. The URL is prone to change. In case of 404, please contact the authors.