XXVII Congresso de Iniciação Científica Unicamp
16 a 18 de outubro de 2019 - Campinas | Brasil

Analysis and Implementation of Algorithms for Similarity Matching of Concepts in Conceptual Maps.

Kaique Chiovetto Siqueira*, Gisele Busichia Baioco.

Abstract

Conceptual maps are graphical tools that represent the relations between concepts. They are useful for learning, assessment, organization and representation of knowledge by assisting in the retention and retrieval of information during the learning process. Therefore, they are used in the most different purposes. However, there are difficulties in comparing conceptual maps referring to the syntax and semantic of the concepts used. This work aimed to implement, analyze and qualify algorithms for similarity comparison of concepts present in conceptual maps.

Key words:

Conceptual Maps, Similarity Comparison, Concepts

Introduction

Conceptual maps are graphical tools used to organize and represent knowledge. They contain propositions that are expressions about some object or event in the universe. Each proposition contains two concepts connected by a connecting word in order to compose a meaningful statement, as described by [1]Novak and Cañas.

[2]Moreira said that there is no correct conceptual map, but a conceptual map according to the meanings that a person attributes to the concepts and the relations between them. Since the author of a conceptual map has a way of representing his knowledge, the syntax and semantic of the concepts become an important factor for concept maps comparison, in order to improve the understanding of a given subject. Therefore, it is necessary to identify, analyze, implement and test algorithms for similarity matching of concepts in conceptual maps.

Results and Discussion

First of all, it is important to understand that concepts are formed by few words that indicate a perception that someone has on a certain subject in a certain scope. Then, considering that concepts are short phrases, we started looking for similarity comparison algorithms of words or phrases.

The first algorithms identified deal only with the syntax of the concepts to perform similarity matching (syntactic similarity). The algorithms Longest Common Subsequence (LCS), Levenshtein, Jaro Winkler, Hamming and Damerau-Levenshtein were implemented and analyzed. It was possible to conclude that Jaro Winkler, Hamming and Damerau algorithms are not indicated to perform concepts comparison. The comparisons using Jaro Winkler algorithm returned false-positive values, Hamming algorithm is not adequate since the concepts compared must have the same number of characters, and the Damerau algorithm is an extension of the Levenshtein, then the results of both are very similar.

After that, we searched for algorithms that deal with the semantic of concepts (semantic similarity). The Jaccard Similarity and the Cosine Similarity algorithms were implemented and analyzed. In order to the concepts are properly manipulated by the algorithms is necessary

preprocessing them. The preprocessing consists of the concepts' segmentation in a set of words, the removal of articles, pronouns and adverbs, and the words are left in lowercase.

Some results of the algorithms are presented in Table 1. We use the concepts in Portuguese language. The algorithms return values between 0 and 1, where 0 means distinct concepts and 1 equal concepts.

Table 1. Results of the comparison of concepts

Concept 1	Concept 2	LCS	Levenshtein	Cosine	Jaccard
experiências	Experiências	0,92	0,92	1	1
Boa experiência	Experiências	0,67	0,6	0	0
Novas experiências	Experiências	0,61	0,61	0,71	0,5

Based on the results, it is possible to conclude that the algorithms based in semantic similarity present more accurate results. However, they need preprocessing, generation of binary vectors with the words of the concepts, beyond the similarity calculation algorithm, making the execution time larger than the syntactic similarity algorithms. Considering the code implementation complexity, Levenshtein and LCS algorithms are easier to implement. They use only substrings and addition, removal and editing techniques.

Conclusions

The accuracy of similarity matching of concepts in conceptual maps depends on concepts syntax and semantic. However, more accurate results increase execution time and complexity of algorithm implementation. Thus, this work aimed to help future works that need conceptual maps comparison in choosing the more adequate algorithm to compare concepts according to each application requirements.

Acknowledgement

[1] NOVAK, J. D.; CAÑAS, A. J. *The Theory Underlying Concept Maps and How to Construct and Use Them*. **2008**.
[2] MOREIRA, M. A. *Mapas Conceituais e Aprendizagem Significativa*. **2012**.