Semantic-Awareness Recommendation with Linked Open Data in Web-Based Investigative Learning

1910155 TING Kang

With the rapid development of information and communication technology in the 21st century, human life has also undergone tremendous improvements. In the field of education, we have considerable expectations and emphasis on the development of the web resources such as Linked Open Data(LOD) which is combined a blend of Linked Data and Open Data. LOD breaks down barriers between different data formats and sources. Web-based investigative learning is one of the learning approaches benefited from.

As the model of Web-based investigative learning provides a platform for learners to create their own learning scenarios by organizing knowledge over the web in self-directed way. This kind of knowledge management activity helps learners to achieve a high cognitive load on investigation. However, it is difficult for learners to discover related concepts among a vast number of unstructured web resources concurrently with a better knowledge construction process. Therefore, this research aims to propose a method to recommend semantic-related concepts with Linked Open Data for learners during the investigation of the web-based investigative learning process.

We proposed a Semantic-awareness Recommendation System which extracting the relevant concepts at different levels from DBpedia. DBpedia is a linked open data project which extracts structured content from Wikipedia. Those structured content represented as Resource description framework(RDF) graph allowed the user to query the relationships and properties of Wikipedia resources semantically. In this work, generating a regulated concept map based on the initial question for the recommendation, three significant elements would be considered:

- Semantic relations: According to the SKOS document, the properties broader and narrower are used to assert a direct hierarchical link between two concepts.
- Node importance: The PageRank algorithm would calculate the importance of the concepts extracted by regulated SPARQL query strategy.
- Content containment: It is based on concept utility. For example, not every concept has definition in DBpedia. The hypothesis is which concept without definition is not important for the recommendation.

For the concepts extraction, we proposed a Regulated Concept Map Generation process by using regulated SPARQL query strategy. We firstly extract Simple Knowledge Organization System(SKOS) Concepts(RDF graph)

from DBpedia using SPARQL query. Then, related concepts with semantic relations(Broader-Narrower) would be returned. The essential property: **SKOS:broader** would be used. This property represents a hierarchical relation between concepts. It is important for us to regulate the SPARQL query strategy if we aim to recommend the related concepts at different levels without preventing learners from their self-directed investigation. The regulated concepts map is a collection of entities called nodes, which are concepts that we are going to recommend to learners. Concepts are linked by edges with the property ***SKOS:broader***(Broader) and ***is SKOS:broader of***(Narrower).

Since the PageRank algorithm is generally used as an index to decide the importance of nodes in a directed graph such as the RDF graph. Therefore, the PageRank algorithm is suitable for the concept importance estimation of this work, and we named it as Semantic-aware PageRank. We assume that the importance of a concept node is determined by the number of outbound links on that concept. The probability of random surfer a node is weighted by the total number of nodes in the Regulated Concept Map.

Before updating the recommendation list to learners, we have to filter those concepts which are not important. In this work, we would filter concepts based on the concept's utility. Since not every concept has a definition on DBpedia, the hypothesis that the concept has no definition on DBpedia is not significant for the recommendation.

For the evaluation, the concept importance estimation results would be analyzed by Spearman's correlation coefficient. Spearman's correlation coefficient measures the strength and direction of the association between two ranked variables. we had compared the strength and direction of the association between DBpagerank, User expectation, and Semantic-aware PageRank by Spearman's correlation coefficient. Owing to the finding of analyzing results, Semantic-aware PageRank maintained most serious strength of the association between User expectation. Furthermore, a case study was conducted for testing the hypothesis that using our proposed recommendation system could help learners strengthen the knowledge construction process by discovering semantic-related concepts during Web-based investigative learning. The results evaluated by statistical methods suggest that Semantic-awareness recommendation with linked open data promotes the efficiency of the knowledge construction process.