DOI: https://doi.org/10.46354/i3m.2019.vare.009

INTELLIGENT NAVIGATION OF LINKED DATA WITH A GRAPHICAL INTERFACE BASED ON SEMANTIC SIMILARITY

Pedro Fonseca-Ortiz[a], Ceballos H. G.[b]

[a],[b], Tecnológico de Monterrey

[a]pedrofonseca@tec.mx, [b]ceballos@tec.mx

ABSTRACT
Semantic Web Technology proposes the use of linked data and ontologies as a mean for providing meaning to information. Even though several tools for the analysis and visualization of linked data exist, these tools require a lot of specialized knowledge to fulfill a purpose. Additionally, this complexity hardens its use for non-experienced users therefore limiting semantic web applications. This paper describes a tool that combines the use of a recommendation system and an intuitive dynamic user interface for navigating linked data. The tool guides the user to find resources of interest by highlighting those related to his search intention. This is, the platform learns on the fly the user interest and makes recommendations based on the connections between resources.

Keywords: Semantic Technology, Linked Data, Recommendation System, Educational Innovation

1. INTRODUCTION

Information is one of the most valuable resources nowadays. The good use and analysis of data are of great benefit for almost every human activity. Eventhough the great amount of existing data and the exponential growth of it has open new uses for the information itself. However, in order to take real advantage of data, it is necessary to have the capability to process, analyze and read it correctly. A proposal for getting the best of data is the use of the Linked Data structure of RDF triplets (Bizer, Heath and T. Berners-Lee 2009). This structure uses the concept of triplets to add meaning to an object in a database. This by connecting each object to other resources in the set using the subject, predicate and object composition established by the World Wide Web Consortium (W3C-REC 1999). This concept applied to web apps is called Semantic Web.

Semantic Web can be described, in concept, as an improvement to the traditional web. This, according to the World Wide Web Consortium. (W3C 2013), by adding meaning to the data in order to create more legible information for computer applications. This is achieved by adding semantic and ontology metadata to the existing data. This additional information helps to describe the content, meaning, and relation of data (Feigenbaum et al. 2008). In the book "Semantic web for the Working Ontologist" (Allemang and Hendler 2008) semantics is described as understanding the nature of meaning. This can describe web semantic as a group of tools that tends to give intelligence to the web, or at least to endow it the capability to understand a more specific meaning of information as a whole and for users as individuals.

Another topic of consideration for semantic data is the representation of information. In the book "Adaptive Semantics Visualization" (Nazemi 2016) it is described how visual representation that adapts to data and human behavior can benefit the user experience and information acquisition. If we want to take full advantage of linked data, it is necessary that users can understand the information that is retrieved by the system. Furthermore, the user needs to be able to interpret the connections between data.

Semantic Web technology is considered as a set of ideas that seek to improve web performance by adding meaning to the resources on the network. However semantic web technology only set the foundations to apply different types of models that take advantage of linked data. This work proposes an algorithm, that in addition to a dynamic visual interface, take advantage of the semantic connections between data to calculate the relevance of an object to a user. This relevance factor is used to make better recommendations to the user with every click.

2. LITERATURE REVIEW

2.1. Semantic Web

Semantic Web technology is a set of concepts and ideas developed upon the World Wide Web platform. The semantic web is developed under the schemes of W3 (World Wide Web Consortium) and is built upon the premise of adding semantic and ontologic metadata inside the administration of databases on the internet (W3C 2013). With this description it can be concluded that Semantic Web is the idea of a web that recovers, access and deploys the data an user needs based on the relations between objects. In 2001 Tim Berners-Lee (Tim Berners-Lee, Hendler and Lasila 2001), known as the father of the world wide web, describe semantic web as a new type of network wich content have a meaning for the computers

2.2. RDF

As described before, the semantic web is a set of activities and elements for the organization, management, and retrieval of data inside an ontology. Given that in semantic web one of the most important concepts is the relation between data, the W3C organization certifies a syntaxes for declaring objects in a database alongside their relations. This scheme is called RDF (Resource Description Framework) (W3C 1999). The RDF scheme is the way information is distributed on the semantic web. The RDF syntax also knows as a triplet, consist of a first element called subject, that is connected to another object (that can also be a subject in another triplet) by a predicate. With the RDF scheme, a database can be visualized as a big graph where all the objects are nodes where the connecting lines are the predicates.

2.3. Linked Open Data

Linked Open data is an architecture that consists of practices within the web to connect and give structure to data. The result of this architecture is a type of database where the objects are connected between them with a semantic relation. This with the objective of making more useful data (Bizer, Heath and Tim Berners-Lee 2009). One of the intentions of the development of linked data is that data is available in a manner that computers can understand it. This with the objective of make more efficient queries. With this the idea of making connections between objects in different databases is possible. According to Tim Berners-Lee (Bizer 2009), with linked open data it is possible to make information "readable" for computers, information with specific and well-defined meaning and have interconnected databases.

2.4. Recommender Systems

A Recommender System is an information filter that based in a model and algorithm identify objects inside a database that has relevance for a specific user. At the same time, a recommender system removes information that is redundant or of no interest for the user. A recommender system compares retrieved information from the user with some characteristics references that can be obtained from the user experience or from a community of similar users (Hanani, Shapira and Shoval 2001). There are two main recommender systems models: based on collaborative systems and based on content.

3. DYNAMIC USER INTERFACE FOR DATA VISUALIZATION

One of the objectives of this research is to apply the concept of recommender systems to take advantage of linked data connections. In this paper an algorithm that shows objects related to each user query is proposed. The algorithm uses a content-based filter to perform recommendations. The recommendations are done based on the relevance of each object connected to the already retrieved objects and the selections of the user. In order to make this a relevance vector is calculated. This vector helps the system to rank the connected objects and recommend the best ranked. This semantic relation between objects helps to guide the user to find better results to a query. This because the algorithm can learn the relevance an object has to the user and adjust this relevance in every click. An important part of this work is the visualization strategy, which the algorithm takes advantage of to make better recommendations. Semantic web technology lays the groundwork for creating visualization adapted to the user, the tasks and the data (Nazemi, Burkhardt, Ginters et. al. 2015). Furthermore, with the application of artificial intelligent techniques, such as the algorithm proposed in this work, it is possible to take advantage of the semantic technology structure to create a more adaptive visualization

3.1. Visualization Strategy

For the visualization strategy, a web application with a dynamic adaptive user interface is proposed. This strategy is focused on showing the related objects to every query result and the highlight of each node. For most relevant objects the nodes will have a larger diameter. This relevance is recalculated in every user selection and therefore the diameter of each node. An example of this is shown in figure 1. The application of adaptive visualization strategies has been used in linked data structure before showing significant benefits to regular visualization strategies (Nazemi, Burkhardt, et. al. 2014). A benefit of dynamic visualizations it that the semantic relations of data can be better represented according to the user's behavior.

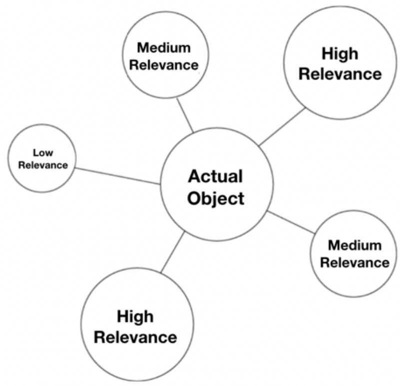

Figure 1: Basic visualization layout

This because even though the objects semantic relations in a database don't change, the relevance of those relations to the user may vary depending on what is the user searching for.

3.2. Cold start

One of the biggest issues of the recommendation systems algorithms is to start from a state where the reference data is small or is none. This issue is called

cold-start problem (Qiu, et. al. 2011). To solve this problem, the *Personalized Video Ranker PVR* system from *Netflix* was used as a reference for this work (Gomez-Uribe and Hunt 2015).
For the case of study, the user is presented with a series of initial recommendations based on the semantic similarities found in the database. For creating this recommendation, a clustering algorithm was used to find the main clusters in the database. The clusters were found considering the tags of the objects and the relations to other similar objects. Once the clusters were identified the object in the center of the group, the one with more connections in the cluster, is taken as the starting point for the search. With this, the cold star problem is reduced. This because even if the user doesn't know the objective resource of the search, at least the user knows the topic. The semantic connections are then used to navigate faster throughout the cluster.

3.3. Data Navigation

Here is specifically described each one of the events that the user will fire to navigate in the proposed recommendation strategy.

1. The user access to the front page of the web app. In this section, the most representative pictures of each category are shown. The user then selects the one that represents better his interests.
2. A first layout is shown with the most representative object of the category as the center.
3. With each click of the user to an object, the relevance to the related objects is calculated. The most relevant objects are shown as the layout shown in figure 1.

One of the main ideas for the visual recommendation navigation is that in a Linked Data architecture all objects are at least connected to another resource in the database. Therefore, while starting the search from an initial object (initial node) to a target object (final node) there is a search space between them.

The recommender system's objective is to find the shortest path. This by searching for the target object by learning those characteristics (connections) that the users favor in each click. This even if the user doesn't know what is the target object. To do this, the system must highlight those objects that represent on higher degree those characteristics the user favors (the most relevant).

Based on this, the expected behavior of the web app is to show, in each iteration, 10 possible nodes to continue the search starting for the latest clicked object. This means that for each iteration the object selected by the user becomes the central node in the layout. From this selected node 10 object recommendations are displayed as nodes considering the parameters of relevance described above. This is repeated in each iteration until the user finds the target object. This process is shown in figure 2

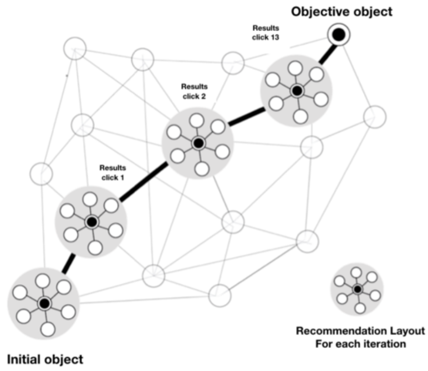

Figure 2: Basic example of the recommendation visual Layout behavior

4. RELEVANCE VECTOR ALGORITHM

The relevance value is the pivotal part for the recommender system to work. In this process, the *weight* with which each object is going to be shown to the user is calculated. It is then deducted that a more relevant object is highlighted more. A characteristic matrix for the relevance of the objects is calculated and the values on it are updated in each iteration. This relevance weight adjustment is considered as learning for the algorithm. This matrix $M \rightarrow R \times F$ describes R RDF resources in terms of the triplets that describe them (F). Below is described each element that composes the recommendation algorithm.

- R : Is the list of RDF resources.
- F : The triplet existence <R predicate-object> is represented with 1 and it absents by 0 in a SPARQL endpoint.
- |R| : Denotates the number of matrix rows.
- |F| : Denotates the number of matrix columns and is given by the quantity of combinations predicate-object for each resource $r \in R$.
- $S \subset R$: Denotates the objects selected by the user (clicks).
- M[r,f] : If the value 1 is shown its means the <r,f> in the SPARQL endpoint exist, and 0 in the opposite case.
- f : Denotates the index of the column related to the relation predicate-object.
- $\vec{s}$: Is a similarity vector of size |F| where $s[i] = \sum_1^n M[r,f]$ such that $r \in S$ and $n = |R|$
- $\vec{e}$: Is the relevance vector with size |R| where $e[i] = \sum_1^m M[r,f] \times s[f]$ where $r \in S$ and M=|R|
- $\vec{e}_n$: Is a normalize vector in scale of [1-10] where 10 represents the higher relevance $\vec{e}_n = e[i]/\max(\vec{e}) \times 10$ (rounded to integer)

- min_viz : Visualization threshold, any object which relevance value is below this value, will not be shown but will remain in M
- $V \subset R$: Indicates which resource $r \in R$ have $e_n[i] > \min \_viz$
- $|V|$: The quantity of resources shown in the visualization.

The table 1 is an example of a characteristic's matrix in which the relevance value is calculated.

Table 1: Example of characteristic's matrix

Characteristic's matrix						
Resources		TOPIC			Relevance	Normalized
		Topic 1	Topic 2	Topic 3		
X	F1	1	1	1	6	100
1	F2	1		1	5	90
2	F3	1		1	5	90
3	F4		1	1	4	70

The figure 3 is an example of similarity vector ($\underset{s}{\rightarrow}$) in which the weight for each relation is calculated based in the number of selected objects that share the relation.

Figure 3: Example of similarity vector

4.1. Relevance vector calculation

The final step for the algorithm is to calculate the relevance value for each resource. This means how much an object is graphically highlighted in the web application. For this, the vector of similarity $\underset{S}{\rightarrow}$ is obtained. From here the relevance value for each object in the list is calculated in the base of the weight of each label it poses. Therefore, a calculation in which each element k is multiplicated by the value of the label i (1 or 0) for the element with the same index inside the vector $\underset{s}{\rightarrow}$ and adding all the results. This considering that all the results will be given by:

$$\sum k_2{}^{i} \underset{s}{\rightarrow}[i] \times A[k,i] \qquad (1)$$

in which those elements that contain the related topic with a higher weight in the matrix will have a bigger relevance value.

The resulting vector $\underset{R}{\rightarrow}$ contains the relevance value for each object in the list. This vector is normalized in a vector $\underset{R}{\rightarrow}!$ and is calculated in the following way:

$$\underset{R}{\rightarrow}! = \frac{\underset{R}{\rightarrow}[i]-(\min \underset{R}{\rightarrow}-10)}{\max \underset{R}{\rightarrow}-(min \underset{R}{\rightarrow}-10)} \times 90 \qquad (1)$$

5. NAVIGATION OPTIMIZATION

Part of this research work is to prove that a recommendation system algorithm improves the efficiency of an object search in a linked database. This can be measure by the number of clicks (selections) c =|S| the user does before finding the objective object. This means that the system objective is to reduce the number of clicks. This leading the user to select objects that are better evaluated (high relevance value)

6. CASE OF STUDY WEB APP FOTOTOTECA

For this case of study, the database *Fototeca* (Fototeca. Tec de Monterrey) from *Patrimonio Cultural del Tec de Monterrey* (Patrimonio Cultural Tec de Monterery 2009) was used. Fototeca is a database which holds historic photos. Fototeca holds over 6 thousand objects with over 900 tags available. The resources of this database are linked in *RDF triplets*. This connected structure fits for testing the proposed recommendation algorithm. Furthermore, the triplets structure can be visually represented in the user interface. The web application work as a recommendation system using a graphic interface in which the connections of the resources are visualized.

For this work, a web application was developed for retrieving resources from a SPARQL endpoint implemented in a Fuseki server. The visualization was developed using D3. The application receives as input the *URI* of the most connected resource of the category chosen by the user. The front page is represented in figure 4.

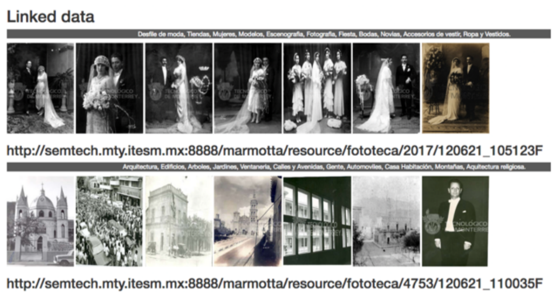

Figure 4: Front page of the web app showing the categories and the most connected phot of each one

Once the user selects an initial object, the system retrieves the 10 most similar photos in the database. The app displays the graph as shown in figure 5. The layout is updated in each iteration where the relevance value is recalculated and represented by the size of the node. The main idea for this is that at each click the algorithm recalculates the relevance value based on the similarity of the pictures. The graphic layout helps the user to visualize the objects that are similar to the actual object and the relevance of each resources to his search.

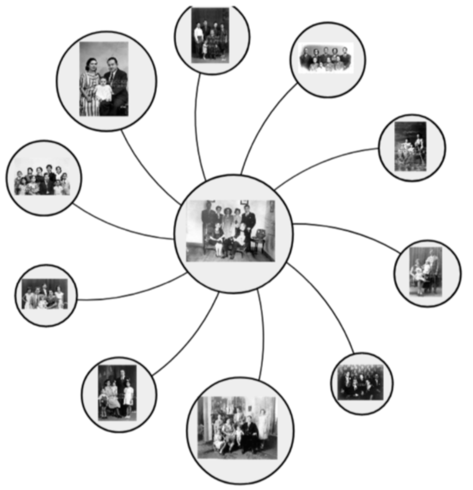

Figure 5: Graphic layout of the recommendation algorithm

7. RESULTS

For the case of study, it was asked to 50 users to search a specific photo using the web app with the recommendation system. The objective of the experimentation was to learn if the recommendation system can guide the users to the objective photo. The results were measured considering if the user finds or not the objective photo, and how many clicks the user makes.

From the 50 users in the test 41 find the objective photo and 9 did not find it (82% find the target photo). For the users that find the target photo the average relevance value for the relevance vector was 70.16 and for those who not find the objective photo was 68.02. This means that those users that follow the recommendations of the system have better chances to find the target photo.

The figure6 shown the average number of clicks in which the user finds the target photo. Also in this figure, it is shown the average relevance value for each click. We can see that users that find the target resource tends to select photos with higher relevance value in each click.

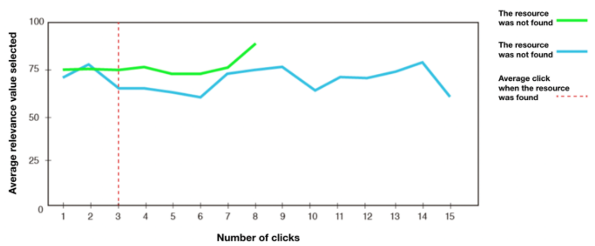

Figure 6: Average value of the relevance vector in. each click

From the 50 users in the test the average age for the users that find the target photo was 22 years and 31 for those who not find the target photo. It was found that younger users have better chances to find the target photo. This can be attributed to the fact that younger users consider more the size of the node as part of a recommendation. Older users tend to consider the content of each photo rather than the recommendation based on the similarity.

The results of this work show that a recommendation system based in similarity to offer relevant resources to a user can be built with Semantic Web technologies and Linked Data. Also, the use of a dynamic adaptative graphic interface to show the connections between the objects and the relevance value helps the user to find the target resources. Also, the selection of the graphic interface helps the algorithm to receive as input the relevant attributes for the user. This work aims toward the optimization of the graphic interface in order to analyze if the selected objects are the best route from the initial object to the target object.

8. CONCLUSSIONS AND FUTURE WORK

This article explores the idea of using recommender systems based on semantic similarity with dynamic visualization in a linked data structure. With this, the intention of taking advantage of semantic relations between resources to calculate the relevance of an object to a user is explored. Based on the case of study results, it can be concluded that dynamic visualization based on relevance in linked data bring considerable benefits to a traditional interface. The concept of adaptative visualization in linked data is a relevant topic in semantic technology. Even though the use of artificial intelligence tools such as recommender systems is still a challenge. This work proves that the use of recommender systems directly connected to a dynamic visualization strategy improves user experience. Future work can be done with the use of emerging technology techniques such as virtual and augmented reality.

An interesting approach for this type of work is the benefit of education innovation. This because one of the most important parts of any educational process is access to information. Furthermore, students and academic professionals must be able to interpret data efficiently. One problem, when retrieving information in most traditional databases, is that data interpretation is subjected to the labels given to each resource. By contrast, in human knowledge semantics is very important to the meaning of information. With tools like the algorithm proposed in this work, it is possible to encourage the use of linked data structure in academic databases. With these, resources like scientific documents can be directly connected by meaning to other objects like authors, subject or institution, not just by labels. Furthermore, tools like these can use semantic connections for helping users to easily discover new resources of great relevance to the query. Future work in this field could include the use of recommendation system algorithms in Open Educational Resources (OER) discovery. Examples of semantic technology apply to OER are Europeana and DBpedia. Using this artificial intelligence tools could enhance user experience and resource discovery.

As today only the relevance value is represented graphically in the interface (node size). Further work in the improvement of graphic visualization is recommended. This to enhance the user experience in order for the user to understand better the semantic relation between resources. Examples of another approach in the graphics interface to be considered in future work is the use of force layout to show the degree of relation between objects and edge weight (connecting line) to represent the similarity between resources.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support of Writing Lab, TecLabs, Tecnologico de Monterrey, Mexico, in the production of this work.

REFERENCES

Allemang, Dean and James Hendler (2008). Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL. San. Francisco, Ca, USA: Morgan Kaufmann Publishers Inc.

Berners-Lee, Tim & Hendler, James & Lassila, Ora. (2001). The Semantic Web: A New Form of Web Content That is Meaningful to Computers Will Unleash a Revolution of New Possibilities. ScientificAmerican.com.

Bizer, C., Heath, T., & Berners-Lee, T. (2011). Linked Data: The Story so Far. In A. Sheth (Ed.), Semantic Services, Interoperability and Web Applications: Emerging Concepts (pp. 205-227). Hershey, PA: IGI Global.

Feigenbaum, L., Herman, I., Hongsermeier, T., Neumann, E., & Stephens, S. (2007). The semantic web in action. Scientific American, 297(6), 90-97.

Carlos A. Gomez-Uribe and Neil Hunt. 2015. The Netflix Recommender System: Algorithms, Business Value, and Innovation. ACM Trans. Manage. Inf. Syst. 6, 4, Article 13 (December 2015), 19 pages.

Hanani, U., Shapira, B., & Shoval, P. (2001). Information Filtering: Overview of Issues, Research and Systems. User Modeling and User-Adapted Interaction, 11, 203-259.

Nazemi, K. (2016). Adaptive semantics visualization (Vol. 646). Basel: Springer.

Kawa Nazemi, Dirk Burkhardt, Egils Ginters, Jorn Kohlhammer, Semantics Visualization – Definition, Approaches and Challenges, Procedia Computer Science, Volume 75, 2015, Pages 75-83.

Nazemi K., Burkhardt D., Retz R., Kuijper A., Kohlhammer J. (2014) Adaptive Visualization of Linked-Data. In: Bebis G. et al. (eds) Advances in Visual Computing. ISVC 2014. Lecture Notes in Computer Science, vol 8888. Springer, Cham

Qiu, T., Chen, G., Zhang, Z. K., & Zhou, T. (2011). An item-oriented recommendation algorithm on cold-start problem. EPL (Europhysics Letters), 95(5), 58003.

(1999). "Resource Description Framework (RDF) Model and Syntax Specification", W3C. URL: http://www.w3.org/TR/PR-rdf-syntax/. [<Accessed 10 / 2017>].

(2013). "W3C Semantic Web Activity, World Wide Web Consortium (W3C)". World Wide Web Consortium (W3C)". [<Accessed 10 / 2017>].

(1999). "Resource Description Frame- work (RDF) Model and Syntax Specification", "W3C-REC". URL: http://www.w3.org/TR/PR-rdf-syntax/. [<Accessed 10 / 2017>].