Recommendation system for the Brazilian IT sector through unstructured text

Patrick Anderson Matias de Araújo[1], Gian Franco Joel Condori Luna[1]

[1]Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Cidade Universitária "Zeferino Vaz" – 13.083-970 – Campinas – SP – Brazil

patrick.araujo@students.unicamp.br, g234826@dac.unicamp.br

Abstract. *It has been detected that in the IT sector there is an entire ecosystem, in which buyers look for sellers that meet their needs. Currently there are many problems when trying to hire companies that provide IT services due to the rudimentary way in which they are currently being carried out, especially in the Brazilian context. One of the problems detected is the poor elaboration of the requirements when developing an IT project, since most of those responsible for the elaboration of the project are company personnel who do not have much knowledge of the subject. It is proposed a recommendation system based on ontologies that serve to help the development of IT projects by providing the most appropriate profiles and technologies for the project.*

1. Introduction

The trade in IT services has grown rapidly creating many new companies, and this growth has become an important part of the economy of many countries, transforming IT into a complete economic sector. It has been detected that in the IT sector there is an entire ecosystem of purchase and sale of products and services, in which those involved try to contact the best option that can successfully meet their requirements.

Currently there are many problems when trying to hire companies that sell products and/or provide IT services, since this search is still carried out in a manual and rudimentary way, based on networks of trust, friendships and word of mouth from collaborating companies ending up with solutions that do not match the specifications of the required problem and causing a loss of time and money for the company.

In the Brazilian IT sector, technological solutions are known, such as web platforms, that try to solve this problem but are not having the expected results, since it was detected that the user companies are not preparing a good project where the problem they present is explained. It was possible to detect that the staff in charge of developing IT projects for companies, are mostly staff from different areas of the company that do not have mastery of the subject, so this can generate a poor definition of the necessary requirements for the understanding of the IT project, ending in associations with IT vendors that are not adequate.

1.1. Objectives

The main objective of the practical project is the development of a recommendation system for the Brazilian IT sector through unstructured text, for which we have to meet the following specific objectives:

- Understand how web platforms help IT buyers and sellers to develop their projects.

- Perform the logic of the recommendation system.
- Elaboration of the ontology.
- Development of the practical project.

1.2. Work phases

Our practical project will have the following phases of development:

- In the first phase, the operation of web platforms for the IT sector will be investigated in order to obtain the information that allows us to have a general idea of how to approach the problem.
- In the second phase we will think about how we are going to adapt the problem detected for an ontology that can, through queries, give recommendations to clients.
- In the third phase we will see what tools we can use and what is the scope that we can have with them.
- In the fourth phase we will develop our ontology embodying in it the operating logic of the IT sector on the web.
- In the fifth phase, the creation of the recommendation system with the selected tools is carried out.

2. Methodology

To solve our problem, we first had to understand the functioning of the web platforms that currently exist to solve the problem of buying and selling IT services and/or products in Brazil. For which we rely on the web platform created by the company "Match IT" that served as a guide for the development of this practical project. We started by creating two users, one an IT buyer and the other an IT seller, to be able to analyze how this platform works and obtain the logic of the business.

Knowing the logic of the business, we began to develop our solution, for which we require tools that can help us. We needed a simple and friendly environment where we could start to develop our ontology, for which we had among the options: "Protect", "Python" and "Java". All allow us to create ontologies. We chose Protect because it allows us to have a visualization of the ontology while we are creating it, something that does not happen with the other two options.

After obtaining our ontology we needed the interaction of the ontology with users, as for example the user should be able to create and query for instances of a certain class, in an easy and transparent way. The best known programming languages that allowed us to have this ability were Python and Java. Both were very good at using ontologies, both allowed for Sparql queries and also had a lot of documentation, but we went with Python as we have more experience working with it.

Finally, within Python we needed to define which package to use to perform the ontology manipulation. We had RDFlib, which manipulates the ontology using RDF, and Owlready2, which uses owl, we tried both, but we liked Owlready2 more because it was more intuitive and simple, especially for queries with Sparql because we could easily provide the necessary parameters for a consultation.

Having already defined everything that we are going to use, we begin by making the ontology. Using the Protect tool we create the following classes, their respective object properties and data properties:

- Empresa: It is the class that will represent each IT company. It has the following Data Properties:
    - "tem_id_empresa": Relationship that points to a Literal that stores the unique identifier of the company.
    - "tem_nome_empresa": Relationship that points to a literal that stores a text that indicates the name of the company.
    - "tem_localizacao_empresa": Relationship that points to a Literal that is the place where the company is located.

  It has two sub classes: "IT_Seller" and "IT_Buyer".

- Vendedor_TI: It is the company in charge of selling IT products and/or services in Brazil. It's a subclass of "Empresa" and inherits the company data properties, so we don't have to redefine them. In addition to the inherited properties, it has the object property
    - "tem_perfil_ti": It is responsible for relating to the class "Perfil_TI", since each seller is related to a certain IT profile and it is through this that they can be linked with a respective IT project and thus locate an IT buyer .
- Comprador_TI: They are the companies that are on the lookout for IT vendors. They inherit the properties of the "Company" data and in addition to this they have the following object properties:
    - "tem_palvra_chave": It will help us to be able to relate the IT buyer with the "Palavra_Chave" class, which is where the instances of all the keywords that the IT buyer has will be stored. It has its symmetric property "belongs_buyer_TI".
    - "tem_projeto": It allows us to relate the "Buyer_IT" class with a project, since each IT buyer has to carry out the elaboration of a project in order to find the most suitable seller.
- Palavra_Chave: It will serve to obtain the most appropriate IT profiles regarding the keywords that are written in the summary field of the IT project. It has the following object properties:
    - pertence_comprador_ti: It will allow us to know to which IT buyer a keyword belongs. It is symmetric of "tem_palavra_chave"
    - "tem_perfil_ti": It is responsible for linking the keyword with its corresponding profile. This will help us later to make recommendations.

  It also has the data properties:

    - "tem_id_palavra_chave": Relationship that points to a Literal that is a unique identifier of each keyword.
    - "tem_nome_palavra_chave": Relationship that points to a literal that is the name of the keyword, it is the text that will be searched for in the project summary so that we can make recommendations.
- Projeto: Represents the project that the buyer needs to develop to explain what their needs are. You have the following object properties:

- “tem_perfil_ti”: Every project must have one or several profiles depending on the problem it is going to address.

It has the following data properties:

- “tem_id_projeto”: Relationship that points to a Literal that stores a unique identifier for each project.
- “tem_nome_projeto”: Relationship that points to a Literal that stores the name of the project.

- Perfil_TI: This is where the instances of the different types of categories or profiles that exist in the information technology sector will be created. It has the following object properties:
  - “tem_tecnologia”: Each IT profile instance has an instance of the “Technology” class that helps solve a specific problem. Then this relationship serves to connect both instances. It is symmetric to the property “pertence_perfil_ti”. It has the following data properties:
  - “tem_id_perfil_ti”: Relationship that points to a Literal that is a unique identifier for the TI profile.
  - “tem_nome_perfil_ti”: Relationship that points to a Literal that is the name of the IT profile. These are the names that will be displayed to the IT buyer when recommendations are displayed.
- Tecnologia: They are computer tools that help in solving the various problems that arise in the IT sector. It has the following object properties:
  - “pertence_perfil_ti”: It is the property that will allow us to get from a “Tecnologia” instance to an instance of “Perfil_TI”. It is symmetric to the property “tem_tecnologia”. It has the following data properties:
  - “tem_id_tecnologia”: Relationship that points to a Literal and stores the unique identifier of the Technology instance.
  - “tem_nome_tecnologia”: Relationship that points to a Literal and stores the name of the Technology.

The Technology class has as a subclass the classes: "Banco_Dados", "Frontend", "Backend" and "Outras".

Having completed our ontology, now it's time to export the OWL file to be able to import it from Python and be able to manipulate the creation of instances and queries in a more practical way.

In the python programming language, functions were created that allow us to create each of the instances of the classes that we have in our ontology and their respective data properties, and the function returns the object that points to the ontology instance.

The functions created are:

- criarEmpresa(nome:str, localizacao:str): Receives two parameters, one of the name of the company to be created and the second parameter is where this company is located.
- criarComprador(nome:str, localizacao:str): Receives two parameters, one of the name of the company to be created and the second parameter is where this company is located.

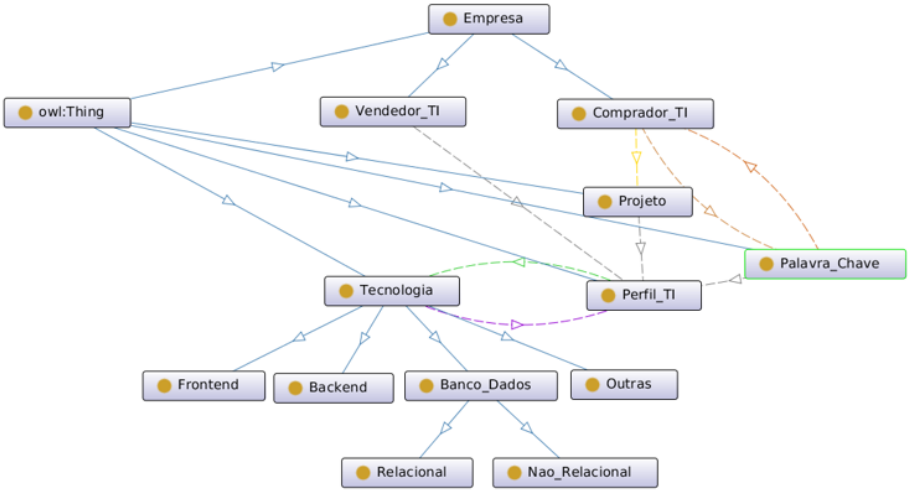
- criarVendedor(nome:str, localizacao:str): Receives two parameters, one of the name of the company to be created and the second parameter is where this company is located.
- criarPalavraChave(nome:str): Receives the name of the keyword.
- criarProjeto(nome:str): Receives the name of the project.
- criarPerfilTI(nome:str): Receives the name of the IT profile.
- criarTecnologia(nome:str): Receives the name of the Technology.
- criarBancoDados(nome:str): Receives the name of the keyword.
- criarRelacional(nome:str): Receives the name of the relational database.
- criarNaoRelacional(nome:str): Receives the name of the non-relational Database.
- criarFrontend(nome:str): Receives the name of the Frontend technology.
- criarBackend(nome:str): Receives the name of the Backend technology.
- criarOutras(nome:str): Receives the name of other technologies.

Then we proceeded to define functions that help us perform the SPARQL queries we need. Among the most complex functions we had the one that received parameters to be able to give a result, such as the function "mostrarRecomendacao(company, array_pc:list)" that received a python object as a parameter that is representing an instance of the Company class of our ontology and also received a list of keywords. Through these keywords, the ontology had to be consulted if the company instance had a keyword instance with the same name and, accordingly, it returned a recommendation on the screen.

To show the recommendation, the “recomendacaoGeral (array_pc_geral)” functions are also used, which is in charge of showing a general recommendation of IT and technology profiles, which are the ones that the web platform administrator defines when creating the words keys, "recomendacaoPersonalizada (company, array_pc_personalizada: list)" which is responsible for searching in the ontology to which IT profile and through this to the technology that the keywords belonging to a certain company belong to and finally the function "recommendationByUser (array_pc:list)” which receives a list of keywords and finds that other users have these same keywords and displays the IT and technology profiles that these other users used.

3. Results

The result of the ontology made in Protege is shown in the following figure:



In order to show the results of the recommendation system we are going to show 3 scenarios:

- First scenario: The objective of this scenario is to be able to see the general recommendations for which an "exemplo 1" buyer is created who is located in "campinas" and who is entering our recommendation system for the first time. When writing the summary of your project, mention the word “servidor”.

```
RECOMENDACOES PERSONALIZADAS
----------------------------
----------------------------
RECOMENDACOES POR USUARIO
----------------------------
----------------------------
RECOMENDACOES GERAIS
--------------------
soluciones cloud computing
- mongodb
- alibaba
- aws
- mysql
- oracle
- azure
soluciones crm
- powerbi
- sap
- oracle
- sas
--------------------
```

Interpreting the results: Our recommendation system shows the personalized recommendations section empty since this company does not have previous projects because it is the first time that it is entering and it does not have recommendations from other users because it is the first user that entered our

systems. What can be seen in the image is the general recommendation, which is the recommendation that shows the IT profiles that are related to the keyword that was entered in the IT project summary.

- Second scenario: The objective of this scenario is to be able to see the recommendations by user for which two users will be entered:
    - Buyer “teste 1” located in “Sao Paulo” who used the keyword “pagina web” in the summary of his project and connected this project with the IT profile “Soluciones de Desenvolvimento de Software” and this IT profile uses the technologies "php", "html" and "mysql".
    - Buyer “teste 2” located in “Minas Gerais” also used the keyword “pagina web” in the summary and also selected the IT profile “Soluciones de Desenvolvimento de Software” but with different technology, he chose “python”.

  Now we enter a third buyer user named "example 2" with location in "barao geral", it is the first time he enters our system. He has the keyword "pagina web".

```
RECOMENDACOES PERSONALIZADAS
----------------------------
----------------------------
RECOMENDACOES POR USUARIO
----------------------------
soluciones de desenvolvimento de software
- php
- mysql
- html
- python
----------------------------
RECOMENDACOES GERAIS
--------------------
soluciones de desenvolvimento de software
- angular
- html
- reactjs
- mysql
- css
- oracle
- c++
- python
- mongodb
- php
--------------------
```

  Interpretation of the result: This result shows us the empty personalized recommendations section since it is the first time that it enters our systems, but compared to the previous example, it shows us recommendations from other users, which were the users previously entered.

- Third scenario: The objective of this scenario is to be able to see the personalized recommendations for which we will use the buyer user already generated in the

second scenario "teste 1" and add the keyword "pagina web" in the project summary.

```
RECOMENDACOES PERSONALIZADAS
----------------------------
soluciones de desenvolvimento de software
- html
- php
- mysql
----------------------------
RECOMENDACOES POR USUARIO
----------------------------
soluciones de desenvolvimento de software
- php
- mysql
- html
- python
----------------------------
RECOMENDACOES GERAIS
--------------------
```

Result interpretation: It shows us in the personalized recommendations section the IT profile and the technologies that the purchasing user had already used in a previous project. It also shows us the recommendations by user. We no longer thought it convenient to show the general recommendations because with two recommendations we believe that it is enough.

4. Discussion

- The tools used were sufficient to achieve our goal.
- A lesson learned may be that you have to understand the problem very well and define the solution well before starting to build the ontology, since you can have changes that can alter, by inheritance, other classes of the ontology.
- We would like to know how we can put our ontology into production.

5. Conclusions

We can conclude that if we have a good dataset that helps us feed our ontology with keywords, IT profiles and technologies used in IT projects in Brazil, we could help improve the development of IT projects by buyers, to achieve optimal search for companies selling services that solve your problems.

6.- References

Ali, Farman, et al. (2019) “Transportation Sentiment Analysis Using Word Embedding and Ontology-Based Topic Modeling”. In: *Knowledge-Based Systems*, vol. 174, page 27–42. https://doi.org/10.1016/j.knosys.2019.02.033.

Awangga, Rolly Maulana, et al. (2019) “Ontology design based on data family planning field officer using OWL and RDF”. In: *TELKOMNIKA (Telecommunication*

Computing Electronics and Control), vol. 17, page 161, https://doi.org/10.12928/telkomnika.v17i1.9237.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001) "The Semantic Web. A New Form of Web Content That Is Meaningful to Computers Will Unleash a Revolution of New Possibilities". In: *Scientific American*, 284, pages 1-5.

Chen, C., Huang, H., Ross, K.E. et al. (2020) Protein ontology on the semantic web for knowledge discovery. In: *Sci Data* 7, 337. https://doi.org/10.1038/s41597-020-00679-9.

Dos Reis, Julio Cesar; Martins, Tulio Brandão Soares. (2022) "Handling Multi-chapter Inconsistencies in DBpedia Evolution". In: *SIMPÓSIO BRASILEIRO DE BANCO DE DADOS (SBBD)*, 37, 2022, Búzios. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, pages 128-137. ISSN 2763-8979. https://doi.org/10.5753/sbbd.2022.224307.

Hitzler, Pascal. (2021) "A Review of the Semantic Web Field". Communications of the ACM, vol. 64, pages 76–83. https://doi.org/10.1145/3397512.

Introduction — Owlready2 0.36 documentation. (2022). https://owlready2.readthedocs.io/en/v0.37/intro.html.

Jain, S., Mehla, S., Agarwal, A.G. (2019). An Ontology Based Earthquake Recommendation System. In: Luhach, A., Singh, D., Hsiung, PA., Hawari, K., Lingras, P., Singh, P. (eds) *Advanced Informatics for Computing Research*. ICAICR 2018. Communications in Computer and Information Science, vol 955. Springer, Singapore. https://doi.org/10.1007/978-981-13-3140-4_30

Jupp, Simon, et al. (2016) "The Cellular Microscopy Phenotype Ontology". In: *Journal of Biomedical Semantics*, vol. 7, page 28. https://doi.org/10.1186/s13326-016-0074-0.

K. U. Danyaro, J. Jaafar and M. S. Liew; (2012) "An RDF model for meteorological and oceanographic information systems". In: *International Conference on Computer & Information Science (ICCIS)*, 2012, page. 480-484, https://doi.org /10.1109/ICCISci.2012.6297293.

Kontopoulos, Efstratios, et al. (2013) "Ontology-Based Sentiment Analysis of Twitter Posts". In: *Expert Systems with Applications*, vol. 40, pages 4065–74. https://doi.org/10.1016/j.eswa.2013.01.001.

Kume, S., Masuya, H., Kataoka, Y., & Kobayashi, N. (2016). Development of an Ontology for an Integrated Image Analysis Platform to enable Global Sharing of Microscopy Imaging Data. In: *ISWC2016 The 15th International Semantic Web Conference*. https://doi.org/10.48550/arXiv.2110.10407.

Michael Uschold (2001). Where are the semantics in the semantic web. In: *AI Magazine*, 24, 2003.

Noy, Natasha, et al. (2019) "Industry-Scale Knowledge Graphs: Lessons and Challenges". In: *Communications of the ACM*, vol. 62, pages 36–43. https://doi.org/10.1145/3331166.

Protégé. (2022). https://protege.stanford.edu/.

Semantic Web - W3C. (2022). https://www.w3.org/standards/semanticweb/.

Match IT (2022) https://matchit.com.br/