Open Government Knowledge Base

Stainam Nogueira Brandao
COPPE/Federal University of Rio de Janeiro
Rio de Janeiro, Brazil
stainam@cos.ufrj.br

Tiago Silva
COPPE/Federal University of Rio de Janeiro
Rio de Janeiro, Brazil
tiagoss@cos.ufrj.br

Sergio Assis Rodrigues
COPPE/Federal University of Rio de Janeiro
Rio de Janeiro, Brazil
sergio@cos.ufrj.br

Luis Araujo
Brazilian Department of the Treasury
Brasilia, Brazil
luis.araujo@planejamento.gov.br

Jano Souza
COPPE/Federal University of Rio de Janeiro
Rio de Janeiro, Brazil
jano@cos.ufrj.br

Abstract — **The paper describes the governmental knowledge base constructed from Brazilian laws, the events that surround them and authorities responsible for these. The knowledge base is built on RDF (Resource Description Framework) language through the Brazilian Official Gazette, which is the access for official information. This knowledge base is linked to other Brazilian Open Data and being one more effort in Open Government Partnership [10] to reflect the country's commitment to strengthen the transparency of government action, to prevent and combat corruption. The self-management characteristic in the knowledge base was obtained through an architecture that monitors the laws status in all Brazilian Open Data Providers and highlights the inconsistencies in a period of time. This system is used by the Secretary of Federal Treasury and presents a well-founded basis for checking the accuracy and validity of the decrees in a Federal Budget domain.**

Keywords-Knowledge Base; Self-Management; Linked Open Data.

I. INTRODUCTION

In the context of Open Data [10], Brazil has been actively participating and the last news was a federal law recently launched for the purpose of free access of federal information by Brazilian citizens [2], taking into account that some information is fundamental to the safety and operation of government public companies.

The main information source is the Federal Official Gazette, which is a PDF document and contains legislation, jurisprudence and administrative actions. Published by authority since 1808, the today's Brazilian Gazette is the Brazilian Government's Official Journal. It was set up to provide King Jonh with news while he and his court were in Brazil publishing Decree, Laws, Program and Internal Rules. With a new edition every day, today's Brazilian Gazette contains a huge amount of information and statutory notices about decisions and changes in a local and national level. The Brazilian Gazette is a natural candidate for the government semantically enables the potential reuse of information.

Our goal is making public sector's information available in a way that enables and maximizes its reuse for citizens, consumers and government adding maximum value. Furthermore, we notice that due to the complexity of managing a large number of documents within an organization, a mechanism for automating the documents annotation task has become essential in order to facilitate the retrieval of information.

Sacramento et al. [22] defines and relates ontology, knowledge base and data sources such as used within this research:

(a) An ontology is a pair O=(V,S) such that

(i) V is a finite alphabet, the vocabulary of O, whose atomic concepts and atomic roles are called the classes and properties of O, respectively, and

(ii) S is a finite set of inclusions in V, constraints of O.

(b) A knowledge base is a triple KB=(V,S,A) such that

(i) (V,S) is an ontology, and

(ii) A is a finite set of assertions in V.

(c) A data source is a pair DS=(V,A) such that

(i) V is a finite alphabet, and

(ii) A is a finite set of assertions in V.

Within this work, ontology, knowledge base and data sources are used with this caution. Similarly, RDF (Resource Description Framework) [7] is a triple subject-property-object, usually described as P (S, O), where a

given subject S has a property P that assumes the value O. Sacramento et al. [22] defines Linked Data as a set of best practices for publishing and connecting structured data on the Web [24]. From the user's perspective, the main goal of Linked Data is the provision of integrated access to data from a wide range of distributed and heterogeneous data sources [25].

The linked data is achieved through interaction with other Brazilian open knowledge bases developed independently, making use of technology patterns according to [17].

The self-management characteristic was incorporate in this work through an application that monitors the RDF (Resource Description Framework) triples and validates the integrating facts between several datasets. In this paper, self-management knowledge base is defined as a knowledge base that monitors and analyses the data linking between the data sources used to.

The aims of this project are two fold:

- Address the practical challenges of publishing public governmental information in a way that maximize its reuse and ensure valid and consistent information;

- Give to Brazilian Gazette a new role, as a vehicle of semantically enabled official information, ensuring publicly available.

II. RELATED WORKS

Our project is inserted in the Brazilian Initiative Linking Open Data for moving the Web from the idea of separated documents to wide information space of components easier to discover, more valuable and easier for people to reuse.

In the context of semantic web, applications can help to produce value-added content by simply treating resources information unstructured. There is a need for readers to know where to get new and pertinent information and how to consume it. The RDF (Resource Description Framework) provides a common framework that allows data to be shared and reused across application, enterprise and community boundaries. Also, it provides a simple data formalism for talking about things, their properties, inter-relationships and categories.

According to Berners-Lee [27], the raw data should be made available as soon as possible. Preferably, it should be put up as Linked Data. Also, it should be linked to other sources, which will allow any data communication to be composed of many mixed vocabularies.

Joshi et al. [30] proposes to querying linked data by using alignments for processing queries, whose constituent data come from heterogeneous sources. They treat alignment between datasets using similarities of ontology instances. This similarity is calculated by names with natural language processing and, the ontologies and datasets used for.

Heim et al. [29] define relationship discovery via the Semantic Web as a highly user-centered process. They identified semantic relation through user´s social bookmarks.

DBPedia [28] is a typical case of a large Linked Dataset, which essentially, makes the content of Wikipedia [31] available in RDF (Resource Description Framework). The importance of DBPedia is not only that it includes Wikipedia data, but also that it incorporates links to other datasets, e.g., to Geonames [33]. Providing links in terms of RDF triples, a user might exploit the knowledge from other datasets when developing an application through of integrating facts from several datasets.

Nevertheless, a large amount of unstructured textual information is supposed to be processed before annotating and validating information that are publicly available. Some definitions found in the literature define semantic annotation as a specific schema to create and use metadata, enabling new methods of access to information [12]. Eller [5] points out that the semantic annotation of documents describes its contents by the association of relevant resources through predicate, also described in the ontology. Eller [5] also developed a system for semantic annotation of documents related to lawsuits using the Smore tool with support of an ontology in RDF (Resource Description Framework) to facilitate the process of semantic annotation of documents for specific areas.

Many services available on the Internet offer generation of semantic information from textual information. We can highlight OpenCalais [15], Zemanta [16], Ontos [11] and TextWise [14]. All these are characterized by enrichment of web pages content with hints (notes) from knowledge bases as DBpedia, GeoNames, IMDB [34] and Musibrainz [35].

OpenCalais [15] is a service provided by Reuters [13], through which it is possible to create semantic content from unstructured source text. The OpenCalais performs natural language processing (English and French) and also use machine learning techniques to define entities in the text.

The entities are divided into three categories:

- Named entities - people, companies, organizations, books, albums, authors, geographic entities;

- Facts - categorized by position, political events, etc;

- Events - sports, change of command, etc;

According to [15], using these information it is possible to build maps (or graphs or networks) linking documents to people, companies, places, and various other entities. OpenCalais uses the microformat language [6] to express its annotations.

Since API proposed by Zemanta [16] provides functionality for generating semantic information from not semantics textual information, Zemanta Service Web returns a set of information (annotations) associated with the text submit for it, such as photos, magazine articles and links from Wikipedia articles, each corresponding to a resource in DBpedia ontology.

Brazilian LexML Project [19] aims to unify, organize and facilitate the access to legislative and legal information made available in digital form by all authorities of Executive, Legislative and Judiciary at the federal, state and municipal levels.

This project produce a controlled vocabulary that plays a key role for organization of legislative and legal information, allowing the integration of this vocabularies used by different governament houses. For example, the user may inform the prefix "urn:lex:br:lei" in the URN (Uniform Resource Name) of Supreme Court judgment. Thus, "urn:lex:br:federal:lei:2008-06-19;11705" is the standard name for the law n°11705 and internationally recognized unambiguously, which ensure a single URN for each piece of legislation. However, one of the major challenges for the legal domain is the retroactivity principle of the law to meet current regulations.

The vocabularies published by LexML use the W3C SKOS (Simple Knowledge Organization System) format [20] and organize the legislative and legal information, allowing, among other things, conducting process patterns of URN (Uniform Resource Name) and the integration of vocabularies used by different Houses. The Document Type Vocabulary available in (LexML 2011) was used in this work. However, another limitation of this project is the absence of Linked Data, for example, when a law cites another law, without use the identifier of this law. What results in nonexistence of Linked Data, not allow the complete reuse.

III. SIOP-LEGIS

SIOP-LEGIS is a Knowledge Organizational System that mines Brazilian Official Gazette daily generating metadata for each document related to Federal Treasury domain. Nowadays, SIOP-LEGIS repository has a collection of more than 80.000 indexed documents since 2008 and that number increases daily [32]. The goal of SIOP-LEGIS system is to use the most accurate official document set wherever possible to aggregate and interlink document that refer to the same resource (Decree, Laws, Program and Internal Rules). Thus, there is a clear need of applying temporal database concepts to RDF (Resource Description Framework) to allow metadata navigation across time.

Initially, a mining module was proposed to automate the indexing process of documents found in Federal Official Gazette that is a PDF document, which contains a domain documents set considered relevant [32]. As shown in Figure 1, despite being a PDF, there is no publicly available metadata or text format from the government press, which makes more difficult to automatically process of reading and indexing the Official Gazette. The technological challenge was to create a mining algorithm that is not tied to the current layout of the Gazette and not being vulnerable to small changes in the style of the Gazette, which is an unlikely situation.
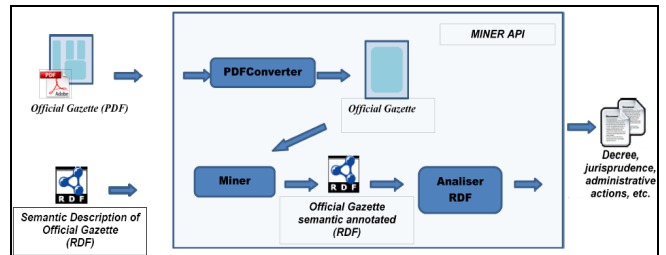

Figure 1. Process of conversion, mining and identification of documents in Federal Official Gazette

For each domain document identified, a metadata is automatically created as shown in Figure 2.

Figure 2. Metadata created automatically (in Portuguese language) for each domain document identified through Dublin Core metadata terms [26]

The mining module was created with the purpose of an API making the whole process transparent to the user, taking as input the Federal Official Gazette (in PDF format) and the RDF (Resource Description Framework) model that specify the structure of Official Gazette. After, the output is the documents list with their metadata. The Figure 3 summarizes the internal process of mining the Official Gazette.

Importantly, the Official Gazette has its text printed in several columns or combination of one, two to three columns.

A. Knowledge representation with RDF ontologies

Our proposal follows the LAG (List of Government Affairs) [9], whose approach is embedded in the e-Ping architecture - Standards of Interoperability for Electronic Government, whose the goal is to define a minimum set of assumptions, policies and specifications that regulate the use of Information Technology in the Brazilian government

domain, establishing the conditions of interaction between federal houses and the society.

B. Resource Recognition

Resources are types of documents found in Official Gazette and already mapped by the Senate through the LexML project [19] within a controlled vocabulary: Decree, Laws, Program and Internal Rules.

The interested resource recognition occurs as drawn on Figure 4, when a resource is crawled on text and its unique identifier is recovered in the third-part LEXML knowledge base as shown in Figure 3, where we can see the URN present in the metadata of a law published by the government.
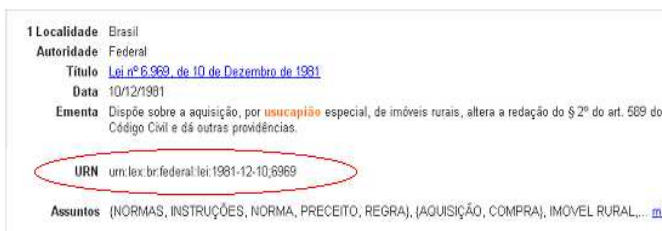

Figure 3. A result search for the law 'Lei nº6.969', which has an unique identifier 'urn:lex:br:federal:lei:1981-12-10-6969'.
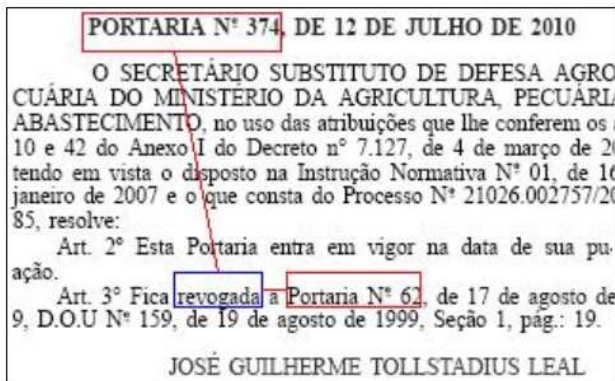

Figure 4. Government action ('revogada', in Portuguese language) identified between two resources: 'Portaria Nº374' repeal 'Portaria Nº62'.

The law cited in the Gazette is mapped in LEXML project to consolidate the Linked Open Data. Right now, the resource has a unique identity URN.

C. Event Recognition

The events between resources are present in the government acts that change the documents status over time: publication, grinding, re-publication, repeal and amendment. Thus, our goal was to process each document and find each governmental act between two resources and represents them on RDF (Resource Description Framework) language [7] to allow, among other things, control the events time.

TABLE 1. EVENTS MODELED BY THE PROPOSED PROCESS

Event	Description	Unique Identifiers
Initiative	Proposition legislative initiative	initiative
Signature	Signature normative documents.	signature
Judgment	Event that will result in a judgment or a decision monocratic.	judgment
Publication	Event official publication of document.	publication
Rectification	Event rectifying an official publication.	rectification
Re-publishing	Event re-publication of official document.	re-publishing
Annulment	Event cancellation of official document.	annulment
Modification	Changing a document on the other, creating a new version.	modification

Table 1 presents the events currently recognized and modeled with their respective unique identifiers. As shown in Figure 4, the events are identified between two resources, thus creating the RDF (Resource Description Framework) triple with the document date. This date will control the chronological order to know the resource status during its cycle life.

D. Authorities Recognition

The proposed governmental knowledge base is also monitoring the signing date, authority and role authority. Through the signing date, the chronological order of decree transformations over time is constructed.
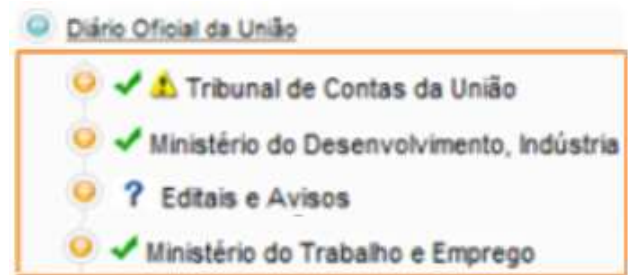

Figure 5. Shows the Brazilian authorities found in one Brazilian Official Gazette

With recognition is possible to identify the authority responsible for the promulgation of a decree. These authorities are in the third-party knowledge base maintend by the Planning Ministry, whose are reponsible for make this information public.

Figure 5 illustrates a document processing of Brazilian authorities identified. An exemple, Court of Auditors (Tribunal de Contas, in Portuguese). We can see three situations: the authority recognized successfully '✓', the authorities recognized with some difference in spelling '✓⚠' and; possible authority, but not recognized by the base '❓'.

E. Named Entities Recognition

Our governmental knowledge base performs the named entity recognition (NER) from other Brazilian open data. A major limitation here is this third-party dataset do not adopted open standard RDF (Resource Description Framework) technology. Even though this limitation, we are

motivated to computationally solve it in order to go forward in our project.

The solution thanks the Brazilian law that obliged all authorities of Executive, Legislative and Judiciary at the federal, state and municipal levels become available public servants with their occupation and function. From this information source, we have developed an algorithm to extract the names and performed the recognition of named entity on the *Corpus*.

Figure 6. An example of third-party official document with possible entities that can sign official decree.

Then, the official document issued by the Ministry, as Figure 6, is processed and used as input for the named entities recognition. In Figure 6, we have two employees with the Unique Personal Identifier (*CPF*, in Portuguese), Employee Name (*Nome do servidor*, in Portuguese) and their actual workplace (*Órgão*, in Portuguese), which are treated as authorities in this paper.

IV. CASE STUDY

Nowadays, our project is used by Federal Budget Secretary in order to audit federal publications within the domain of Federal Budget. The strength of the approach is the ability to audit documents through a user friendly interface that allows users to search for documents related to one of the available parameters as shows in the Figure 7:

- Authority (Autoridade, in Portuguese): grace the named entity identified in the processing;
- Date Signature (Data de Assinatura, in Portuguese) and Publication Date (Data de Publicação, in Portuguese) that used concomitantly with Revoked field (Revogado, in Portuguese) to identify the status of a Document Type on a specific period of time.

Figure 7. A user friendly interface that allows users to search for documents related to one of the available parameters

The processing result can be viewed in the search performed by the graphic user interface for documents or government actions available to the citizens in Portuguese language. Figure 8 shows a search result with RDF (Resource Description Framework) link and the URN of each document highlighted in red, listing with the differential:

- Shows the URN of each document, as present in this paper;
- Shows the possibility of export the RDF document representation, which allows a future automatic processing of documents returned in the search.

Figure 8. Search result with RDF link and the URN of each document highlighted in red

In Figure 9, we can see the RDF document that brings temporal processing of a federal document. From this point is possible to know the document status in a given month/year and the changes ocurrs so far. The temporal question is addressed by the knowledge base with information about each interesting event of selected document (resource).

Figure 9. RDF representation of the government act with the use of controlled vocabularies

The weakness remains the effective natural language processing to identify named entity of interest. Figure 10 shows the needed association between person's name (*Jose Guilherme Tollstadius Leal*, in Portuguese) with his title (in Portuguese, *Secretario Substituto de Defesa Agrope...*), which requires a high precision and recall.
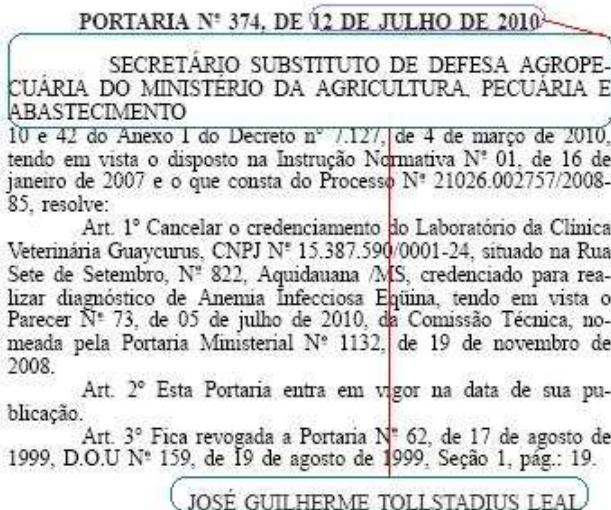

Figure 10. The desirable association between person's name (*JOSÉ...LEAL* , in Portuguese) with his title (in Portuguese, *SECRETÁRIO...ABASTECIMENTO*)

We could have this information in a relational database, but the goal here is the open access to government governamental linked knowledge base in a reuse form and understandable way. Besides creating the base interconnected with other databases that follow the open Brazilian international standard defined by [21].

V. CONCLUSION AND FUTURE WORK

The SIOP project [32] is part of Presidency initiative and Federal Budget Secretary for information provision for society. In addition to providing linked and interoperable open knowledge base, the proposal allows the development of tools to read the data provided by own government.

Note that this need arose for approval of government's public accounts. Moreover, there is a whole temporal control that lets us flag automatically when a rule is valid before causing inconsistency in the knowledge base.

The link between resources and third-party LEXML base, as well as government actions were done without problems because they are defined in a controlled vocabulary with a federal initiative. The big problem that remains is the named people recognition in the document. Despite being included as a future work, the first experiments showed an accuracy and recall of 64% and 61% respectively, which is still a unsatisfactory desire for the project.

Initially, we are working on fully automated process of documents selection in a domain of interest through the creation of rules that contain the terms of a domain ontology. The goal is to extend the application to other domains that has interest in other documents that would be selected automatically for processing.

Another work in parallel is to verify the person responsible for the decree, validating his authority for decree's promulgation in a period of time. The goal is to find inconsistencies in the knowledge base and indicates them for the domain specialist. With this, self-management characteristic will be added to the knowledge base.

Another ongoing future work is the creation of a unique identifier for authorities and people identified through recognition. However, this are suppose be provided by third-party governmental institution responsible for this information, with already happening with the project that provides a LEXML identifier to government actions promulgated, for example, the identifier '*urn: lex: br : ministerio.planejamento.orcamento.gestao: decree :2010-04-28, 777* ' for 'Decree No. 7778 of 04/28/2012', as Figure 7. The current recommended best practice for RDF identifiers is to distinguish between resource identifiers and document URLs. Resource identifiers are used as identifiers within RDF graphs, while document URLs are used on the web to retrieve documents [7].

Finally, a next work is become our Linked Open Government Knowledge Base available in web-service access to meet the information with capable of inference.

ACKNOWLEDGMENTS

REFERENCES

[1]   K. Bontcheva, H. Cunningham, V. Tablan, D. Maynard, and O. Hamza, 2002. "Using GATE as an Environment for Teaching NLP". Proc. D. Radev & C. Brew, orgs. Effective Tools and Methodologies for Teaching NLP and CL. New Brunswick, New Jersey: Association for Computational Linguistics, pp. 53–61.

[2]   General Control of the State, 2012. Information Access Law. Available at: http://www.cgu.gov.br/acessoainformacaogov/ [Retrieve: September, 2012].

[3]   H. Boley, M. Kifer, P. Patrânjan, and A. Polleres, 2007. "Rule Interchange on the Web". Eds. Springer Berlin Heidelberg, pp. 269-

309. Available at: http://rd.springer.com/chapter/10.1007/978-3-540-74615-7_5. [Retrieved: November, 2012].

[4] J. Dolby, A. Fokoue, A. Kalyanpur, E. Schonberg, and K. Srinivas, 2009. "Extracting Enterprise Vocabularies Using Linked Open Data". Proc. 8th International Semantic Web Conference (ISWC2009). Chantilly, VA, USA: Springer-Verlag Berlin, Heidelberg, pp. 779–794.

[5] M. Eller, 2008. "Anotações Semânticas de Fontes de Dados Heterogêneas". Ph.D. Dissertation. Florianópolis, Santa Catarina - Brazil: Federal University of Santa Catarina.

[6] R. Khare and T. Çelik, 2006. "Microformats: a pragmatic path to the semantic web". Proc. 15th international conference on World Wide Web. WWW. New York, NY, USA: ACM, pp. 865–866.

[7] F. Manola and E. Miller, 2004. "RDF Primer". Available at: http://www.w3.org/TR/2004/REC-rdf-primer-20040210/. [Retrieved: February, 2012].

[8] M. R. Koivunen and J. Kahan, 2001. "Annotea: an open RDF infrastructure for shared Web annotations". Proc. of the 10th International Conference on World Wide Web. Hong Kong, Hong Kong, pp. 623-632.

[9] Ministry of Planning, 2011. "Documento da e-PING: Governo Eletrônico". Available at: www.governoeletronico.gov.br/.../lista-de-assuntos-do-governo-lag. [Retrieved: October, 2012]

[10] OGP, 2012. Open Government Partnership. Available at: http://www.opengovpartnership.org/countries/brazil. [Retrieved: October, 2012].

[11] ONTOS, 2010. Ontos Semantic Technologies. Available at: <http://www.ontos.com/o_eng/index.php?cs=1>. [Retrieved: October, 2012]

[12] B. Popov, A. Kiryakov, A. Kirilov, D. Manov, and O. Miroslav, 2003. "Semantic Annotation Platform". Proc. 2 nd International Semantic Web Conference (ISWC2003). Florida, USA Springer, pp. 834–849.

[13] T. Reuters, 2012. Thomson Reuters. Available at: http://thomsonreuters.com/. [Retrieved: October, 2012]

[14] Textwise, 2010. API Overview. Available at: <http://textwise.com/>. [Retrieved: October, 2012]

[15] R. Thomas, 2010. OpenCalais Documentation. Available at: <http://www.opencalais.com/documentation/opencalais-documentation>. [Retrieved: October, 2012]

[16] A. Tori, 2009. Zemanta service. Available at: <http://developer.zemanta.com/files/Zemanta_API_companion_200 80610.pdf>. [Retrieved: November, 2012]

[17] GLD, 2012. Government Linked Data Working Group. Available at: http://www.w3.org/2011/gld/charter. [Retrieved: October, 2012]

[18] K. Bontcheva, H. Cunningham, V. Tablan, D. Maynard, and O. Hamza,, 2002. "Using GATE as an Environment for Teaching NLP". Proc. D. Radev & C. Brew, orgs. Effective Tools and Methodologies for Teaching NLP and CL. New Brunswick, New Jersey: Association for Computational Linguistics, pp. 53–61.

[19] Brazilian Federal Senate, 2011. LEXML: Controlled Vocabularies. Available at: http://projeto.lexml.gov.br/documentacao/ . [Retrieved: October, 2012]

[20] W3Schools, 2011. SKOS: Simple Knowledge Organization System Namespace Document. Available at: http://www.w3.org/2004/02/skos/core. [Retrieved: October, 2012].

[21] W3C, 2012. The World Wide Web Consortium (W3C). Available at: http://www.w3.org/. [Retrieved: October, 2012]

[22] E.R. Sacramento, M.A. Casanova, K.K. Breitman, A.L Furtado, J.A.F. Macedo, and V.M.P. Vidal, 2012. "Dealing with Inconsistencies in Linked Data Mashups". Proc. 16th International Database Engineering & Applications Sysmposium. IDEAS '12. New York, NY, USA: ACM, pp. 175–180.

[23] M.C Pattuelli, 2011. "Mapping People-centered Properties for Linked Open Data". Eds. Knowledge organization. Nord American Symposium on Knowledge Organization. Ergon, pp. 352-359. Available at: http://cat.inist.fr/?aModele=afficheN&cpsidt=24459265. [Retrieved: October, 2012]

[24] C. Bizer, R. Cyganiak, and T. Heath, 2007. "How to Publish Linked Data on the Web". Available at: http://www4.wiwiss.fuberlin.de/bizer/pub/LinkedDataTutorial/. [Retrieved: October, 2012].

[25] C. Bizer,T Heath, and T. Berners-Lee, 2009. "Linked Data – The Story So Far". International Journal on Semantic Web and Information Systems, 5(3), pp.1–22, doi:10.4018/jswis.2009081901.

[26] DCMI, 2012. Dublin Core Metadata Initiative. Available at: http://dublincore.org/

[27] T. Berners-Lee, 2009. Putting Government Data online. Available at: http://www.w3.org/DesignIssues/GovData.html. [Retrieved: October, 2012]

[28] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, S. Hellmann,, 2009. "DBpedia - A crystallization point for the Web of Data". Journal Web Semantics: Science, Services and Agents on the World Wide Web. Volume 7 Issue 3, September, 2009. Pages 154-165. doi: 10.1016/j.websem.2009.07.002.

[29] P. Heim, S. Lohmann, and T. Stegemann, 2010. "Interactive Relationship Discovery via the Semantic Web". Proc. 7th International Conference on The Semantic Web: Research and Applications - Volume Part I. ESWC'10. Berlin, Heidelberg: Springer-Verlag, pp. 303–317. Available at: http://dx.doi.org/10.1007/978-3-642-13486-9_21.

[30] A.K Joshi, P. Jain, P. Hitzler, P.Z. Yeh, K. Verna, A.P. Sheth, M. Damova, 2012. "Alignment-based Querying of Linked Open Data". Proc. Ontologies, DataBases, and Applications of Semantics (ODBASE) 2012. Rome, Italy. pp. 807-824.

[31] D.N. Milne, I.H. Witten, and D.M. Nichols,, 2007. A knowledge-based search engine powered by wikipedia. Proc. Sixteenth ACM Conference on Conference on Information and Knowledge management (CIKM). New York, NY, USA: ACM, pp. 445–454.

[32] S.Brandao, T.S Silva, S.A Rodrigues, L.A. Araujo, D.A. Silva, and J.M. Souza, 2011. "SIOP-LEGIS: Thesaurus for Selection and Management of Brazilian Treasury Domain". Proc. International Conference on Knowledge Management and Information Sharing, Paris, França. pp. *195-200*.

[33] Geonames, 2012. GeoNames Ontology. Available at: www.geonames.org/. [Retrieved: November, 2012]

[34] IMDB, 2012. The Internet Movie Database. Available at: <www.imdb.com/>.[Retrieved: November, 2012]

[35] Musibrainz, 2012. The Open Music Encyclopedia. Available at: <musicbrainz.org/>.[Retrieved: November, 2012]