Integrating a Metadata Editor with Hyperbolic Tree to Improve Data Access in Spatial Data Infrastructures

Marcos Vinícius Montanari, Alexandra Moreira, Vitor Eduardo Concesso Dias, Eduardo Lourenço, and Jugurta Lisboa-Filho(✉)

Department of Informatics, Federal University of Viçosa, Viçosa, MG, Brazil
{marcos.montanari,vitor.dias,eduardo.lourenco,jugurta}@ufv.br,
xandramoreira@yahoo.com.br

Abstract. The storage and management of metadata is an essential task in systems that deal with spatial data. However, to accomplish this task it is necessary to count with a set of tools that operate in integrated mode. This paper describes the development of edpMGB, a metadata editor for the MGB profile. The editor is open-source and is being released in the cloud via Web following the Software as a Service (SaaS) model, so it can be accessed from any site on the Internet, requiring only a Web browser. The editor was developed on the scope of the research and development project (R&D) Geoportal Cemig SDI-based corporate GIS (Geoportal Cemig - SIG corporativo baseado em IDE), whose objective is the implementation of a corporate SDI for the Minas Gerais Power Company (Companhia Energética de Minas Gerais - Cemig). Furthermore, this work proposes the use of a hyperbolic tree to aid indexing metadata, facilitating its recovery. After indexing, the user can navigate through the tree nodes and perform searches over the metadata related to the search words.

Keywords: Spatial Data Infrastructure · INDE · Editor · Metadata · MGB profile

1 Introduction

Within Information and Communication Technology, information redundancy and lack of data standardization proves very common. The same data often ends up being produced, managed, used, and stored by several independent producers that use different formats and standards, who seek to meet exclusively the individual needs of specific users [8].

The greater production of spatial data requires documenting them so they can be reused. A piece of data immersed in its context becomes information, however, with no such documentation, it is virtually worthless information [22].

In order to prevent actions in duplicity and wasted resources to obtain spatial data, the Brazilian government began, in 2003, studies aiming to integrate and

C. Grueau et al. (Eds.): GISTAM 2016, CCIS 741, pp. 15–31, 2017.
DOI: 10.1007/978-3-319-62618-5_2

reuse geo-spatial data produced by the different federal administration organs. In 2008, Act 6,666 of November 27th established the National Spatial Data Infrastructure (Infraestrutura Nacional de Dados Espaciais - INDE) [3]. INDEs goal is “to catalog, integrate, and harmonize the geospatial data produced and maintained by the different governmental institutions so as to facilitate their location, exploration, and access by any user connected to the Internet” [7].

To [18], geographic metadata correspond to the documentation of geographic data and are created according to standards. Such standards consist of a set of regulations that allow the geographic data to be described textually in a previously established manner. The most well-known geographic metadata standards were defined by the Federal Geographic Data Committee (FGDC) and by the International Organization of Standards (ISO). These institutions established international geographic metadata standards that meet the needs of different users, thus which comprises the variability in geographic information [25].

In order to meet the metadata standardization demands started by INDE, the National Cartography Committee (CONCAR) created the Brazilian Geospatial Metadata Profile (MGB Profile) based on norm ISO 19115:2003. A metadata profile is a basic set of elements that portray the characteristics of geospatial products of a given community and guarantees their identification [15].

Based on norm ISO 19139:2007, XML schemas were defined to materialize and code ISO 19115:2003 as a file. Since the MGB profile is based on norm ISO 19115:2003, its metadata must also follow the materialization standards defined in norm 19139:2007 to increase the interoperability among the systems that use the profile as their base [24].

The present paper aims to describe the development of edpMGB, a metadata editor for the MGB profile. edpMGB is open-source software and is available on the cloud via the web following the model Software as a Service (SaaS). Thus, it can be accessed from anywhere and requires only a web browser [30]. Some features of this metadata editor are also available through web services, which allow other systems to use some of this tools features. The editor was developed in the context of the research and development (R&D) project “GeoPortal Cemig SIG corporativo baseado em IDE” (Cemig GeoPortal SDI-based corporate GIS) being developed to help implement a corporate SDI for Minas Gerais Power Company - Cemig (Compania Energética de Minas Gerais).

The main reason for developing edpMGB was the need for a specific tool for this metadata standard focusing on the Brazilian technical audience that works with spatial information. Other editors are available in the market, such as CatMDEdit [6] and Geonetwork [13]. Although these are metadata editors, they are not specific for documenting metadata in the MGB profile. Geonetwork, in particular, can be used alongside the editor described in the present study.

The metadata registered by the editor certainly have a position in the conceptual network underlying the application domain. In order to establish this position, it can use some mechanism to browse the network domain related words. Here we propose the use of Hyperbolic Tree [17] on a network domain words to assist in indexing the registered metadata.

2 Theoretical Framework

This section describes the theories, techniques and tools on which this research sought support, starting on the description of the concepts of Spatial Data Infrastructure and metadata for later addressing standards and related techniques.

2.1 Spatial Data Infrastructure

According to [22], the term "Spatial Data Infrastructure" (SDI) refers to the collection of relevant technologies, policies, and institutional arrangements that facilitate providing and accessing spatial data. It provides a base for discovering geographic data, besides the evaluation and application for users and service providers at all governmental levels, of the commercial sector, of non-profit organizations, universities, and citizens as a whole. Moreover, it hosts the geographic data, attributes, documentation (metadata), and some methods to access such data. A functional SDI must include the organizational agreements required to coordinate and administer it at a local, regional, national, and/or global level in order to enable the ideal environment to interconnect applications to data.

According to [26], the main components of an SDI are:

- Institutional platform related with the policies and administrative agreements in the implementation of standards and data;
- Technical standards that define the technical characteristics of the main data;
- Access network, which make the data accessible by users;
- Data produced in the institutional platform and that meet the technical standards;
- People, comprising users, data producers, and any agent that adds value to the SDI development process.

Figure 1 shows the relation among the elements of an SDI considering people and data as one category and access networks, policy, and technical norms as another. The relation among the categories is very dynamic given the changes in the communities and their needs, as well as the requirement of different datasets.

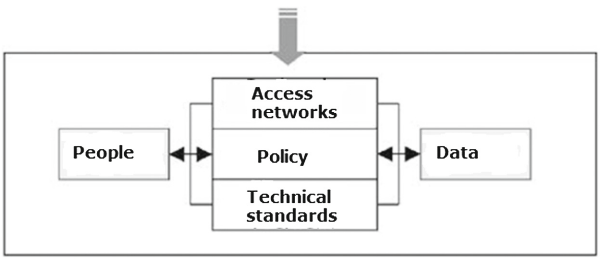

Fig. 1. Nature and relations among the components of an SDI [24].

The importance of the SDI is translated as the potential for sharing geographic information that integrates several types of users. It aims to decrease the lack of interoperability among different systems by standardizing the data. It is a simple concept, but its implementation is hard in view of the technological, political, and cultural differences among the geographic data producers [20].

2.2 Metadata

The increasing production of spatial data furthers the need to document such data so they can be used in future studies. A piece of data with its context (documentation and metadata) becomes information, whereas, without such details, it is virtually worthless information [22].

Metadata represent the information on a piece of data and can be exemplified by descriptions about who, how, when, and why the data was produced. According to Decree-Law no. 6,666 of November 27th, 2008 enacted by the Brazilian Federal Executive Branch, metadata represents the: "Set of descriptive information on the data, including the characteristics of its surveying, production, quality, and storage structure, essential to promote its documentation, integration, and availability, as well as enable searching and browsing them." Metadata use is related to the needs an organization has of better knowing its stored data in more detail. Data cataloging will facilitate its use and meet the users needs. Without this efficient documentation, the users have more difficulty in locating the data required for their applications. This form of data representation aims to contribute to orienting, developing, and describing electronic documents by creating standards, production, and handling in the description by metadata [27].

According to [3], each organ and entity of the Brazilian federal government must share the geospatial metadata they produce. To this end, standards must be used when creating metadata since they are being generated by different organs, however, following the same guidelines to ensure interoperability.

Therefore, the National Cartography Commission (Comissão National de Cartografia CONCAR) defined in 2009 the Geospatial Metadata Profile of Brazil (Perfil de Metadados Geoespaciais do Brasil MGB Profile) in order to establish a national standard used to create geospatial metadata. The profile was developed by several organs that produce geospatial data in Brazil based on international norm ISO 19115:2003. However, this profile does not define how these metadata are structured in electronic files, which is managed by norm ISO 19139:2007, which aims to materialize the concepts of ISO 19115:2003 into a file by coding these metadata in XML schemas. In order to create the MGB profile, some established profiles, also based on norm ISO 19115, were analyzed: MIG Metadados de Informação Geográfica (Portugal); NEM Núcleo Español de Metadados (Spain); NAP North American Profile (EUA/Canadá); LAMP Latin American Metadata Profile (proposed for Latin America) [24].

Norm ISO 19115:2003 used the Unified Modeling Language (UML) to represent the metadata sections. It featured around 400 elements, eight of which mandatory, for profiles derived from this standard. This norm allows defining profiles and extensions for specific application fields [7].

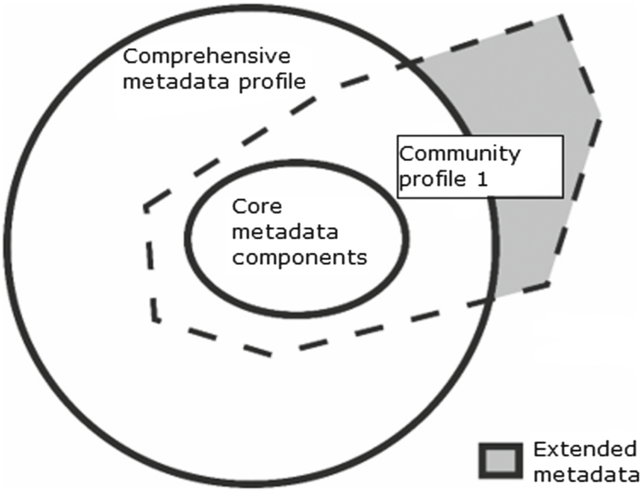

Fig. 2. Metadata profile of a community [6].

A metadata profile is a set of elements that meets the needs of a given community and guarantees their identification. A metadata set defined by the standard may not be enough to meet the needs of a given community, hence the profile may allow additional elements to be used to fulfill the users needs. Figure 2 shows the relation between the metadata set by the standard, the core components, the profile of a community, and the extension made to this profile [7].

The MGB profile has a version called "summarized MGB profile", which represents the minimum set of elements the geospatial metadata produced must have [7]. The abridged profile has 23 elements, presented in Table 1. The user has the option of choosing between the complete version, with 82 elements, and the abridged version. The elements in the MGB profile are organized into the following sections:

- Section 1 (Identification) - provides basic information on the geospatial dataset such as title, date, person or institution responsible, and summary of the data;
- Section 2 (Identification of the geographic dataset GDS) - refers to the information required to identify and evaluate a GDS. This section characterizes the type of spatial representation, scale, language, extension, etc.;
- Section 3 (Restriction information) - publicizes information regarding access and use restrictions and is made up of two entities, one regarding legal restrictions and the other, security restrictions;
- Section 4 (Quality) - allows the quality of a dataset to be evaluated by informing the hierarchical level, linage, and report on the data;
- Section 5 (Maintenance information) - informs the maintenance and update frequency;

Table 1. Entities and elements of the metadata core of the summarized MGB profile [6].

Entity/Element	Condition	Entity/Element	Condition
1. Title	Mandatory	13. Reference system	mandatory
2. Date	Mandatory	14. Linage	optional
3. Responsible	Mandatory	15. Online access	Optional
4. Geographic extension	Conditional	16. Metadata identifier	Optional
5. Language	Mandatory	17. Standard metadata name	Optional
6. Character encoding	Conditional	18. Metadata norm version	Optional
7. Thematic category	Mandatory	19. Metadata language	Conditional
8. Spatial resolution	Optional	20. Metadata character encoding	Conditional
9. Summary	Mandatory	21. Responsible for the metadata	Mandatory
10. Distribution format	Mandatory	22. Metadata date	Mandatory
11. Time and altimetry extension	Optional	23. Status	Mandatory
12. Type of spatial representation	Optional		

– Section 6 (Spatial representation information) - describes the mechanisms used to represent the spatial information (matrix or vector);
– Section 7 (Reference system) - information on the reference system, including the coordinates system and the geodesic referential of the spatial dataset;
– Section 8 (Content information) - describes the catalog of features and the content of the matrix data;
– Section 9 (Distribution) - reports information related to the distributor and to the alternatives to obtain geographic data;
– Section 10 (Metadata) - section responsible for information on its own metadata. It includes the person or institution responsible, creation date, norm used, etc.

Norm ISO 19139:2007 defines a set of XML schemas for metadata defined in ISO 19115:2003. It aims to define a file format for geospatial metadata that follows ISO 19115:2003. These schemas enable structuring and validating the metadata's XML files in accordance with the norm [16]. The XML format allows the metadata instances to go around the Internet, including geospatial web services, as predicted in the specification of the Web Services Catalog [23].

Maintained since 1998 by the W3C, XML is a flexible and simple mark-up language. Its main characteristics are being a text-based language, separating content from formatting, being simple and easily interpreted, and allowing the creation of limitless tags, which facilitates online data exchange [14]. An XML document can be considered well formatted if it matches what it prescribed in the norms [4]. This document can be valid as long as it follows some norms described in its grammar. The XML Schema Definition (XSD) is an XML grammar format [10]. ISO 19139 is described as XML schemas built under the specification by [28].

2.3 Recovering and Visualizing Information

For [29], in communication, an emitter uses mechanisms of information to act aiming to transmit knowledge, which, in turn, must be absorbed by a receiver.

This relation is only possible by using language. Thus, the study of the recovery process involves knowledge in logic, technology, and linguistics.

As information production increased over the years, means were developed to facilitate the process of recovering information to aid the needs of users of both traditional and digital libraries. One such technique to aid recovering information is the use of indexes as a collection of content that indicate where the information the user seeks can be found. Such content can be organized to facilitate search [2]. An example of the importance and use of indexes are in content searches in encyclopedias in which indexes are used to direct the user to the content of their interest without having to read or access other material content.

As the amount of information grows, the complexity of the objects stored also grew and the large volume of data began requiring increasingly enhanced recovery techniques. In face of this new reality, the information recovery process becomes increasingly more important [5].

For [2], the indexing process pertains to the creation of data structures associated with the textual portion of documents such as suffix array and inverted file structures, for example. Those structures may contain data on the characteristics of the terms in the document collection such as the number of times each term appears in a given document.

Information recovery systems present the content found in the corpus and allow the items that meet the users needs to be chosen using a search expression. A simplified representation of the information recovery process is shown in Fig. 3.

Fig. 3. Representation of the information recovery process [11].

Within the information recovery process is the search function, which compares the search expression provided by the users and the documents that are part of the corpus to return the items that feature the information the user seeks. However, even if a term in the search expression features in the representation of a document, that does not mean this document is useful to the user [11].

In this process of information retrieval, data visualization can help in the search carried out by the user. Visualizing information consists in using computing resources to help analyze and understand a dataset. This help is mostly provided by using visual forms [12].

This type of visualization studies the ways of transforming abstract data into real or mentally visible images that facilitate understanding them and helps obtain new pieces of information contained in the data. This process aids in understanding a subject that would be hard to understand otherwise [21].

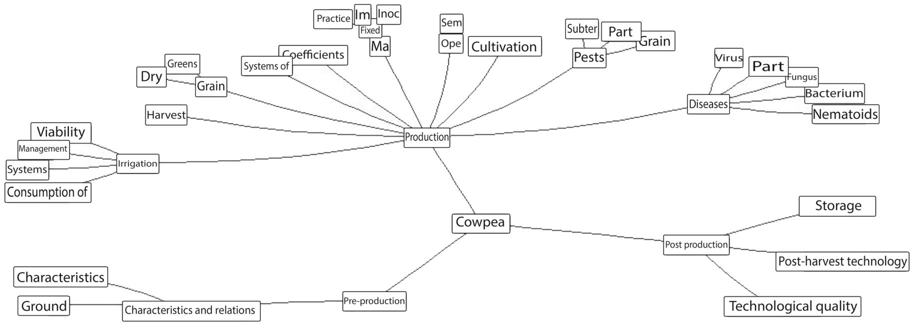

Fig. 4. Example of hyperbolic tree [13].

The visualization techniques represent data graphically to facilitate understanding them. Such representations can be split into three classes: unidimensional, bidimensional, or tridimensional, according to the space dimension where the geometric elements used are located [12].

2.4 Hyperbolic Tree

The hyperbolic tree (Fig. 4) is a graphical visualization resource that facilitates navigation in very large data hierarchies. In this type of representation, the base is the visual form of a tree structure within a hyperbolic space bounded by a circular region. The center of the circle features the observation focus and, the greater the distance from the center, the larger the number of pieces of information presented, while the regions close to the edge of the circle have high information density [12].

This technique is widely used in sitemaps and as a navigation tool in several hierarchies. Links to other pages can be assigned to its nodes. Node color can also be customized [9].

2.5 Cloud Computing and Web Services

According to [1], cloud computing refers to applications provided as online services as well as the hardware and software systems that provide such services. These systems favor the user's access to different applications using the Internet, regardless of the platform used or where the user is. Therefore, all data processing and storage occurs at some online datacenter, as virtual servers.

Cloud computing takes place through different types of services. The main one, which will be approached in this study, is the concept of Software as a Service (SaaS) [31]. This type of service has several resources that the end user can access via web browsers [1]. The metadata editor presented in this study is an example of SaaS.

According to [30], a web service is a software system designed to support the interoperability among machines on a network. A web service is accessed through platform-independent protocols and data formats such as the Hypertext Transfer Protocol (HTTP), eXtensible Markup Language (XML), and Simple Object Access Protocol (SOAP). A web service's interface is accessible through standardized XML messages, i.e., in text format [19].

The editor hereby presented makes available web services that can be accessed via SOAP by any application regardless of the technology used, with no need to access the system's interface. The services can be used by developers that want to use them in their applications.

3 Related Works

Filling out metadata is hard work, as is any other product cataloging.

The use of software to generate geographic metadata files in several formats that meet different norms has been proposed to improve geographic data documentation. Among these tools, the freeware ones such as CatMDEdit and Geonetwork stand out, both featuring forms to fill out the metadata according to a predefined profile and enabling the automatic extraction of some characteristics from the dataset.

3.1 CatMDEdit

CatMDEdit [6] is a tool to edit and visualize metadata in several standards that facilitates resource documentation, particularly geographic information. It is developed by the Instituto Geográfico Nacional de España (IGN) along with the Advanced Information Systems Group (IAAA) from the University of Zaragoza, with technical support from the GeoSpatiumLab (GSL). The tool was implemented in Java and has important features for metadata documentation, among which: The system is multi-platform (running on Linux and Windows); multilingual (Spanish, English, French, German, Polish, Portuguese, and Czech); open source, supporting the automated metadata file extraction and generation (Shapefile, DGN, ECW, FICC, GeoTIFF, GIF/GFW, JPG/JGW, PNG/PGW); and it converts and personalizes metadata standards to generate new metadata standards and profiles in order to serve all types of geographic data.

3.2 GeoNetwork

Geonetwork [14] is a standardized decentralized environment based on a catalog system to facilitate geospatial data access, recovery, update, and management. It provides a complete environment with metadata editor and catalog with search functions. It also carries an online interactive map viewer using web map service. It is currently used in countless SDI initiatives worldwide. Some of its main features include: (1) native support to the metadata standards ISO 19115, ISO 19139, FGDC, and Dublin Core, besides being able do configure a new metadata standard profile; (2) metadata synchronization among distributed catalogs; (3) user management and customized access control; (4) cataloging and access to several types of data and documents (upload/download); (5) interface with multilingual support; and (6) metadata importing in the ISO 19115 standard into a metadata profile configured in Geonetwork. This system is free and open-source, which facilitates its evolution and customization by SDI developers.

What differentiates edpMGB from the editors mentioned in this section is that the former is a SaaS system that does not need to be installed on the users machine, besides having a simplified interface that helps users document metadata. Moreover, it is the only editor specific for the MGB profile and its interface has the local advantage of being in Portuguese.

4 edpMGB - MGB Profile Metadata Editor

Metadata documentation in the MGB profile is an important task for them to be shared and reused. Up until now, no other editor specific for the MGC profile has been created, hence, each metadata set is specified according to the preferences of their authors [24].

edpMGB is a web application developed with the Google Web toolkit (GWT), a framework developed by Google for web-applications. The GWT used the java programming language to develop applications and the tool itself converts the JavaScript code so that the application can be interpreted by any web browser regardless of the platform the user is running.

Through edpMGB, the user can create, edit, and save metadata as a .xml file following international standards so it can be used in several geospatial tools. Before the XML file is generated, it must be validated according to the MGB profile rules. One of the components of GeoPortal Cemig is the geospatial metadata catalog. Therefore, the metadata are documented through edpMGB, thus integrating SDI-Cemig to the INDE.

Figure 5 illustrates edpMGBs home screen. The left-hand side has the navigation tree separated into sections and profile elements. The center-right area features the screen with the MGB profile fields divided into ten panels that represent the sections, which may be accessed using the green arrows to the right or to the left.

Figure 6(a) shows the rules of enforcement, occurrence, and type of value of an element. The icon shown in Fig. 6(b) shows an elements detailed information. The bottom part of the home screen features the editor buttons panel. The button "Abrir" (Open) displays the dialog box where the user can load an XML file to be edited. Each element of the profile will be loaded in its respective text box.

The button "Limpar" (Clear) clears the open elements text box, while "Limpar Tudo" (Clear All) clears all text boxes of the metadata being edited.

The button "Validar" (Validate) performs one of the system's main features, which is to validate whether the metadata is in accordance with the XML schema of the MGB profile. When the button is clicked, the system verifies the data input and may display, for instance, a dialog box as shown in Fig. 7 to inform that the metadata does not respect the MGB profiles rules. The dialog box displays an

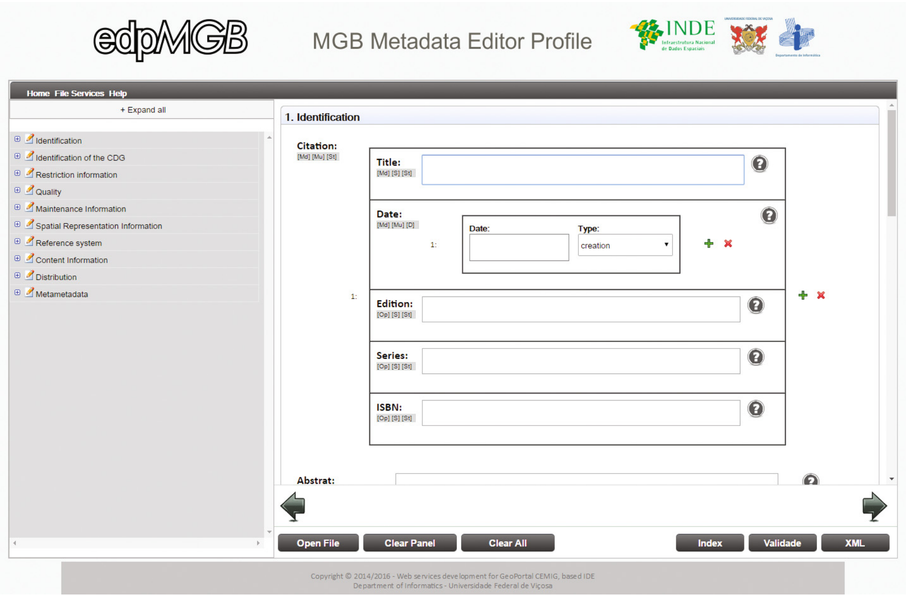

Fig. 5. edpMGBs home screen. (Color figure online)

Fig. 6. Warning dialog about mandatory elements.

error log alerting the user and showing which fields are not in accordance with the rules. The user may keep editing the metadata or store it in his or her machine even if it is not validated for the MGB profile. When a non-validated metadata is generated, it is tagged informing it does not conform to the MGB profile.

A `generateXMLScript()` method receives as parameter the elements of the MGB profile, which are validated by the method `validateMGB()`, then the former outputs the XML script as a string.

The method `validateMGB()` validates the metadata according to the MGB profile, receiving its elements as parameter and outputting a list with the error messages found in the metadata or an empty list in case it respects the profile.

To [24], most metadata provided by national data producers do not fully respect the profile's rules, which is a big issue since it compromises the interoperability among the systems that use the same profile. Nonetheless, the impossibility of saving a non-validated metadata may cause problems to users, perhaps due to the lack of information on the metadata elements. Hence, the user has the

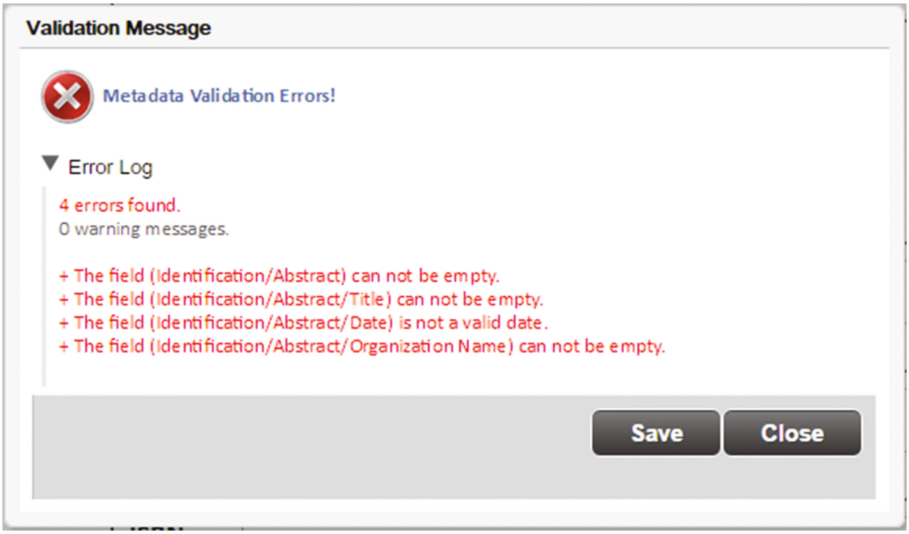

Fig. 7. Result with errors found by the validation service.

option of saving the metadata even if it does not conform to the MGB profile and, if needed, load it again in the tool for further editing. The “XML” button saves the metadata as a .xml file (Fig. 8) that contains all information input in the fields. The .xml file is saved in the users computer.

5 Hyperbolic Tree for the Electric Sector

The edpMGB editor has a function to index the metadata produced and validated by it through a network of words displayed as a hyperbolic tree created for the electricity sector. The Sect. 5.1 explains how the network of words displayed by the hyperbolic tree was built and the way the metadata is indexed is explained in the Subsect. 5.2.

5.1 Extracting Terms Used in the Electric Sector

The terms used to create the hyperbolic tree were extracted using the tool EχATOLP Extrator Automático de Termos para Ontologias em Língua Portuguesa, developed to extract terms from linguistically annotated corpora [19]. EχATOLP is a tool that takes a corpus and automatically extracts all nominal syntagmas (NS) in the text while classifying them according to the number of words. The syntagmas extracted are saved in lists that may have both the NSs in their original form in the text and in its canonical form. The tool also provides some common options for manipulating lists of terms such as applying cut-off points, comparing lists, and calculating usual measures of precision and scope [19].

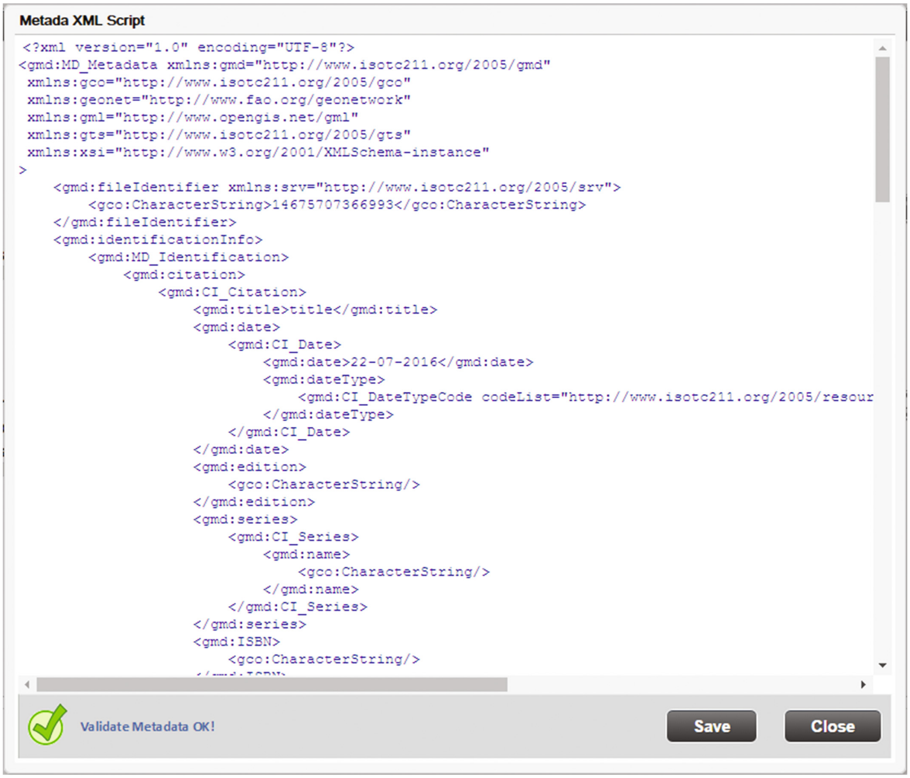

Fig. 8. XML file generation screen.

A domain corpus is a set of texts on a specific field of knowledge that may be used to characterize this domain. In this case, the test base used was the glossary of the National Electrical Energy Agency (Agência Nacional de Energia Elétrica ANEEL). Similarly to other glossaries, ANEEL's glossary is a list in alphabetical order that provides a description of the meaning of words.

Soon after the list of words was generated, the hyperbolic tree was created using Hyper Tree Studio, an open-source tool based on the Hyperbolic Tree library. Figure 9 shows a hyperbolic tree whose central element is the word "Electrical Sector".

5.2 Indexing Metadata in the Hyperbolic Tree

When the Index button is clicked (Fig. 10), the editor verifies whether the metadata are validated and then stores them. After the metadata are stored, the editor creates an address that indicates where such metadata can be accessed in the future. With the creation of this address, the actual indexing process begins.

The editor uses the words typed in the field "título" of the metadata and right after that searches the nodes on the hyperbolic tree. When a node is found

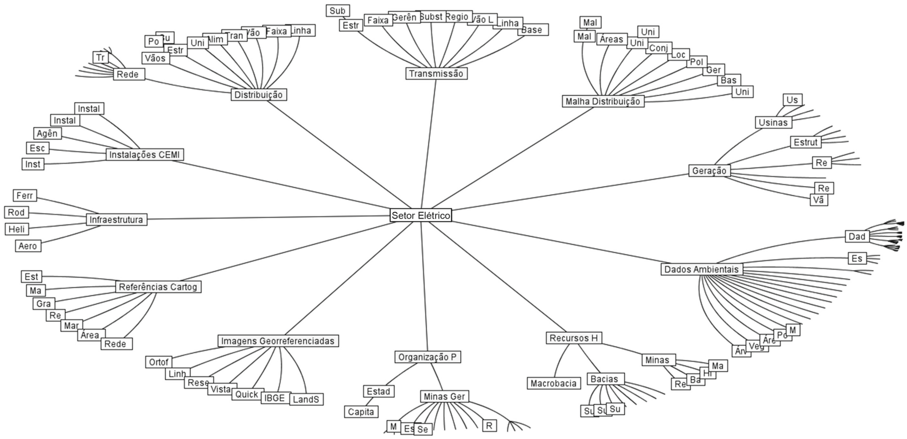

Fig. 9. Hyperbolic tree of the electric system.

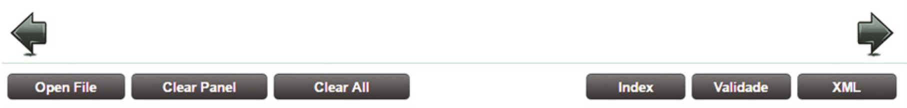

Fig. 10. Panel of buttons in the edpMGB editor.

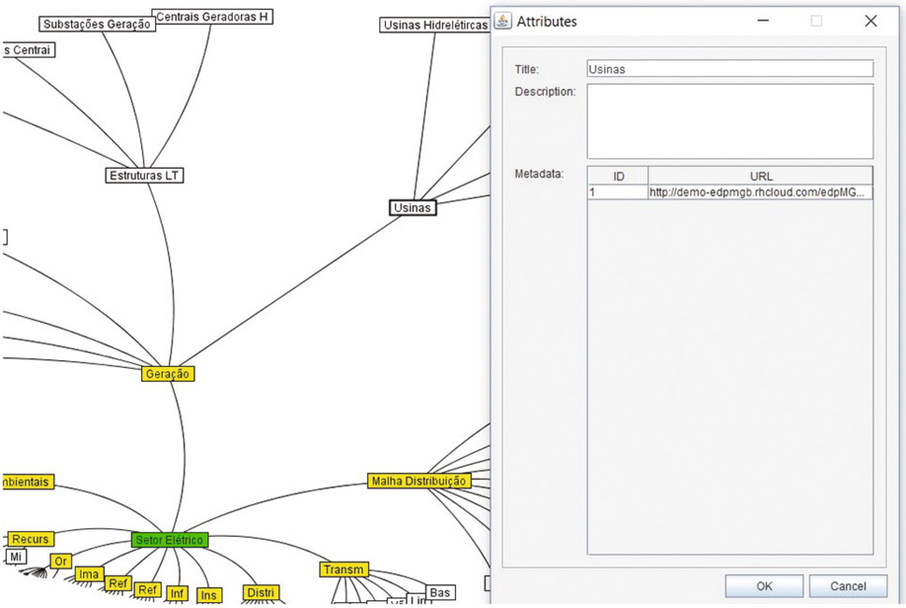

Fig. 11. Attributes box.

that has one or more words in the field “título”, the editor stores within this node the address where the metadata is stored. Finally, the editor exports this tree in XML format with the metadata addresses stored. Figure 11 exemplifies this process. For example, when the node with the word “Usinas” is accessed, the user will find in the window Attributes the field Metadata, which stores the link for the metadata related to this specific term.

6 Conclusions

The present paper presented the development of a geographic metadata editor that follows the concept of Software as a Service (SaaS) and is available to any user with access to the Internet. edpMGB was developed specifically to create geospatial metadata in accordance with the Perfil de Metadados Geoespaciais do Brasil (MGB profile). Besides using software developed in his or her native language, the user has easy access to the system with no need to install it locally. edpMGB may also be used to change metadata sets created with other editors (e.g., Geonetwork).

The feature of validating whether the metadata is in accordance with the standard defined by the MGB profile helps produce higher quality, more complete and correct metadata. However, the user may save the incomplete documents, in which case the metadata receives a tag of non-conformity with the MGB profile. Since the XML validation and generation were also developed as web services, other developers will be able to remotely create applications (e.g., metadata catalog manager) that use the services implemented in the tool through SOAP, which broadens the tool’s reach.

As been shown by [24], most metadata sets currently available in the INDE do not conform to the MGB profile. Therefore, this metadata editor with its XML schema conformity validation service is an important contribution to INDE’s evolution.

This editor was developed in the context of the research and development (R&D) project “GeoPortal Cemig” being developed to help implement a corporate SDI for Cemig - Minas Gerais Power Company. One of the components of GeoPortal Cemig is the geospatial metadata catalog, whose metadata are documented via edpMGB, which integrates SDI-Cemig to the INDE.

It was also described the indexing module attached to the edpMGB editor that allows recovering metadata from a SDI organized within the nodes of the hyperbolic tree. The hyperbolic tree found to be adequate since it allows easy navigation in a complex network of words. The viability of the developed system was verified by developing a case study in the context of a SDI for the electricity sector, using the CEMIG database. It should be noted that the network of words used in the indexing process is a network of links that do not have a formally defined semantics, as well it was not used standardized norms to build the vocabulary. In the future, the development of a domain ontology is needed to improve the indexing process.

Finally, being free open-source software, edpMGB may also be adapted to other geospatial metadata standards and/or profiles.

As a suggestion for future works, the expansion and evolution of the editors specifications is indicated so that an ever more complete mechanism for creation, editing, and search mechanism is provided in the MGB profile. Moreover, further improving the information recovery process in a metadata catalog within the electric sector. Another extension is integrating the editor with automated extraction modules using the bounding rectangle of geospatial data and the treatment of strongly related metadata collections. Finally creating a domain ontology is key for improving the indexing process.

Acknowledgements. This project was partially funded by the Brazilian research promotion agencies Fapemig and CAPES, along with Companhia Energética de Minas Gerais - Cemig.

References

1. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., et al.: A view of cloud computing. Commun. ACM **53**(4), 50–58 (2010)
2. Baeza-Yates, R., Frakes, W.B.: Information Retrieval: Data Structures & Algorithms. Prentice Hall, Upper Saddle River (1992)
3. Brasil: Decreto presidencial n 6.666, de 27 de novembro de 2008 (2008). http://www.planalto.gov.br/ccivil_03_/Ato2007-2010/2008/Decreto/D6666.htm
4. Bray, T., Paoli, J., Sperberg-McQueen, C., Maler, E., Yergeau, F., Cowan, J.: Extensible markup language (xml) 1.1-w3c recommendation 16-august-2006. World Wide Web Consortium (2006). http://www.w3.org/TR/xml11/
5. Cardoso, O.N.P.: Recuperação de informação. INFOCOMP J. Comput. Sci. **2**(1), 33–38 (2004)
6. CatMDEdit: Catmdedit opensource project (2012). http:catmdedit.sourceforge.net/index.html
7. Comissão Nacional de Cartografia - CONCAR: Perfil de metadados geoespaciais do brasil (perfil mgb) (2009). http://www.concar.ibge.gov.br/arquivo/Perfil_MGB_Final_v1_homologado.pdf
8. Dornelles, M.A., Iescheck, A.L.: Análise da aplicabilidade da infraestrutura nacional de dados espaciais (inde) para dados vetoriais em escalas grandes. Bol. Ciênc. Geod., Sec. Artigos **19**(4), 667–686 (2013)
9. Evangelista, S.R.M.: Manual do hipereditor e hipernavegador. Embrapa Informática Agropecuária Documentos (2007)
10. Fallside, D.C., Walmsley, P.: XML schema part 0: primer second edition. W3C recommendation 16 (2004)
11. Ferneda, E.: Recuperação de informação: análise sobre a contribuição da ciência de computação para a ciência da informação. Ph.D. thesis, Escola de Comunicação e Artes Ciências Humanas, Universidade de São Paulo, São Paulo (2003)
12. Freitas, C.M.D.S., Chubachi, O.M., Luzzardi, P.R.G., Cava, R.A.: Introdução à visualização de informações. Revista de informática teórica e aplicada **8**(2), 143–158 (2001)
13. Geonetwork: GeoNetwork User Manual: Release 2.6.4. Geonetwork (2012). http://geonetwork-opensource.org/manuals/2.6.4/eng/users/GeoNetworkUserManual.pdf

14. Goldberg, K.H.: XML Guia prático visual. Alta Books, Rio de Janeiro (2009)
15. ISO: ISO 19115:2003. Geographic information Metadata. International Organization for Standardization (ISO) (2003)
16. ISO: ISO. ISO 19139:2007. Geographic information - Metadata: XML schema implementation. International Organization for Standardization (ISO) (2007)
17. Lamping, J., Rao, R.: The hyperbolic browser: a focus+context technique for visualizing large hierarchies. J. Vis. Lang. Comput. **7**(1), 33–55 (1996)
18. Leme, L.A.P.P.: Uma arquitetura de software para catalogação automática de dados geográficos. Ph.D. thesis, PUC-Rio, Rio de Janeiro (2006)
19. Lopes, L., Fernandes, P., Vieira, R., Fedrizzi, G.: Eχatolp-an automatic tool for term extraction from portuguese language corpora. In: Proceedings of the 4th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics (LTC). pp. 427–431. Faculty of Mathematics and Computer Science of Adam Mickiewicz University (2009)
20. Nakamura, E.T., Queiroz Filho, A.P.: Infraestrutura de dados espaciais: Exemplo do parque estadual de intervales-sp. Revista Brasileira de Cartografia **3**(64), 723–735 (2013)
21. Nascimento, H.A., Ferreira, C.B.: Visualização de informações-uma abordagem prática. In: XXV Congresso da Sociedade Brasileira de Computação, XXIV JAI. UNISINOS, São Leopoldo-RS (2005)
22. Nebert, D.: Developing spatial data infrastructures: The sdi cookbook v. 2.0, global spatial data infrastructure (gsdi). Open Eospatial Consortium (2004)
23. Nebert, D., Whiteside, A., Vretanos, P.: Opengis catalogue services specification (2007)
24. Pascoal, A.P., de Carvalho, R.B., de Araújo Xavier, E.M.: Materialização do perfil de metadados geoespaciais do brasil em esquema xml derivado da iso 19139. In: 16th Simpósio Brasileiro de Sensoriamento Remoto -SBSR. INPE, Foz do iguaçu -PR (2013)
25. Prado, B.R., Hayakawa, E.H., de Castilho Bertani, T., da Silva, G.B.S., Pereira, G., Shimabukuro, Y.E.: Padrões para metadados geográficos digitais: modelo iso 19115: 2003 e modelo fgdc. Revista Brasileira de Cartografia 1(62) (2010)
26. Rajabifard, A., Williamson, I.P.: Spatial data infrastructures: concept, sdi hierarchy and future directions. In: GEOMATICS'80 Conference. Tehran, Iran (2001)
27. Souza, T.B., Catarino, M.E., Santos, P.C.: Metadados: catalogando dados na internet. Transinformação 9(2) (2012)
28. Thompson, H.S., Beech, D., Maloney, M., Mendelsohn, N.: Xml schema part 1: structures second edition. W3C Recommendation, 28 October 2004. http://www.w3.org/TR
29. Vieira, S.B.: La recuperación automática de información jurídica: metodología de análisis lógico-sintáctico para la lengua portuguesa. Ph.D. thesis, Universidad Complutense de Madrid, Madrid (1994)
30. Weiss, A.: Computing in the clouds. Computing 16 (2007)
31. Yang, C., Goodchild, M., Huang, Q., Nebert, D., Raskin, R., Xu, Y., Bambacus, M., Fay, D.: Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing? Int. J. Digital Earth **4**(4), 305–329 (2011)