OLIMPO SYSTEM WEB-TECNOLOGY FOR ELECTRONIC GOVERNMENT AND WORLD PEACE

Hugo C. Hoeschl, Tânia Cristina D. Bueno, Vânia Barcellos, Andre Bortolon, Eduardo S. Mattos

E-Gov, Juridical Intelligence and Systems Institute – Ijuris

Keywords: UN Security Council´s Resolutions, Dynamically Contextualised Knowledge Representation (DCKR), Structured Contextual Search – SCS, Information of Technology, Data retrieve, JAVA

Abstract: The paper describes the Olimpo System, a knowledge-based system that enables the user to access textual files and to retrieve information that is similar to the search context described by the user in natural language. The paper is focused on the innovation recently implemented on the system and its new features. A detailed description is presented about the search level and the similarity metrics used by the system. The methodology applied to the Olimpo system emphasises the use of information retrieval methods combined with the Artificial Intelligence technique named SCS (Structured Contextual Search).

1 INTRODUCTION

Some complex and specific domains require an information retrieval system that is more than just a great technology to search for documents in large text databases. A good knowledge representation is also required.

The present approach enables to retrieve textual information that is similar to the search text described by the user using natural language. Through the extraction of relevant information using DCKR technology (Dynamically Contextualised Knowledge Representation) [8] [9], new documents are automatically included in the knowledge database. Concepts of Case-Based Reasoning (CBR) [1] [2] and information retrieval techniques were applied to obtain a better performance of the system, leading to the technology named Structured Contextual Search – SCS.

The following item 2 of this paper addresses the UN Security Council and the Resolution document; in items 3 and 4 the knowledge representation methodology is presented and Olimpo system is described; in items 5 and 6, describe how the web-technology will be applied to Olimpo system and the impacts that influence not only in the technological field, but also in relation to the citizenship, the knowledge and the research and item 7 is the conclusion of the paper.

2 ABOUT THE UN SECURITY COUNCIL

The importance of the UN body becomes noticeable when one follows the main global means of communication and no further arguments are required. Being the source of the documents handled by the Olimpo system, it is useful to give more details about the Security Council and its document base.

According to its Charter (Article 7-1), the United Nations Organization (UNO) is comprised of six special bodies. All of them issue relevant documentation and it is highly important to have an adequate tool to retrieve those documents.

Given its characteristics and aspects related to the Resolutions, the Security Council was chosen as application field of the Olimpo system.

The Security Council is described by Article 7-1 of UNO's Charter, where it is referred to as a special body of the United Nations. The Security Council is specifically addressed in Chapter V, Articles 23 to 32. As per Article 24-1, its central function is to assume "*the main responsibility in maintaining international peace and security*."

It should be emphasized that the Security Council has a juridical and an executive profile. According to Kelsen (apud Steinfus [11]), it is juridical because it holds the monopoly of legitimate

C. Hoeschl H., Cristina D. Bueno T., Barcellos V., Bortolon A. and S. Mattos E. (2004).
OLIMPO SYSTEM WEB-TECNOLOGY FOR ELECTRONIC GOVERNMENT AND WORLD PEACE.
In *Proceedings of the Sixth International Conference on Enterprise Information Systems*, pages 289-296
DOI: 10.5220/0002644902890296
Copyright © SciTePress

violence at the international scope and judges the existence of facts, determines sanctions on them and who will enforce these sanctions. That turns it a juridical body. And this profile enables a good application of the technology of juridical information, especially SCS and its particular method of rhetoric structure analysis of a given jurisdictional context, based on the knowledge structure involving the body, which maximizes the task of intelligent retrieval of documents when adequate modelling is used.

The Security Council has also political characteristics and it has discritionary power to establish violations, according to Steinfus [11]; therefore the Security Council holds an executive characteristic, turning it a juridical-executive body.

The Security Council presents some peculiarities. One of them is to be currently the most powerful jurisdictional body on the planet. Another one is the existence of internal, informal instances, named "P 3" (Western permanent member countries) and "P 5" (all permanent member countries), according to Steinfus [11]. Another peculiarity is the existence of internal bodies with specific power delegation to perform certain tasks, on a permanent or "ad hoc" level, like the sanctions committee, as shown on Figure 1.

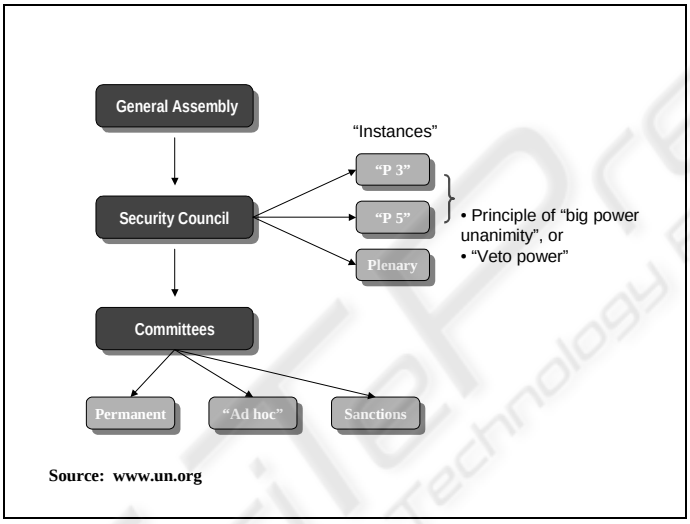

Figure 1: Position of the Security Council

Among the documents issued by the Security Council, six of them have greater relevance. Based on their structure and relevance, the Resolutions were chosen for the application of the Olimpo system.

As per the structure of the document, the Resolutions have some characteristics that make it easier to apply the technology referred to herein.

3 DYNAMICALLY CONTEXTUALISED KNOWLEDGE REPRESENTATION

Olimpo´s performance is centred about the combination of aspects derived from CBR and text information retrieval, in addition to an adequate organization of the knowledge related to the subject the system is focused on (in the present case, the UN Security Council´s Resolutions). The aforementioned knowledge organization is what enables the DCKR technology, which is a methodology that provides the possibility of comparing the contexts described in the documents and not only a comparison between words or attributes.

3.1 Analysis of the rhetoric structure

The rhetoric structure of the system is comprised of indicative expressions used for comparison means and it was, first time ever, dynamically prepared. Up to then it was usual to choose a list of index pointers from a source external to the research group (for example, Court library indexes). Little work was done on the list of index pointers and its selection was based on its similarity with the context of the system under development. For the Olimpo system it was decided to build a particular and specific list, which should be aligned with the issues effectively

treated by the Resolutions and, on the other hand, should be coherent with the documentation context of the managing entity of the database. In this view, in order to collect a list of expressions a detailed reading of the Resolutions was performed, searching onto UNO´s database on the Internet was done and debating with research groups was used. Those expressions were then tested and subject to statistic analysis in order to evaluate their function as reference elements for the indexation and retrieval of documents. A set of expressions with high significance was selected, eliminating those ones with very high or very low frequency of occurrence because they were not very much helpful for establishing a context.

This process had a dynamic characteristic because it was done several times and expressions were included or excluded according to their statistic performance. The routine described shows how it worked to obtain a final list containing a set of expressions that could efficiently reflect the generic, rhetoric structure of the Resolutions, which gave the material form to the dynamically contextualised knowledge representation.

3.2 Structured contextual search – SCS

The searching process being described is said to be "contextual" and "structured" due to the following reasons:

a) For building the rhetoric structure of the system, it is taken into consideration the context of the stored documents;
b) This context is the basis for the input adjustment process, as well as for the comparison and selection of documents;
c) When writing the search text, the input is not limited to a set of words or attributes, but it can take the format of a long text, including the possibility of setting specific attributes, which work as filters and function as a preliminary selection of documents to be searched.

Information contained in the documents is represented in the form of a case, consisting of the original document and a set of eight indexes in the form of pairs of attribute-value: subject, date, number of the Resolution, meeting, country, acronyms, decisions, and indicative expressions. These indexes are part of the system interface (see Figure2).

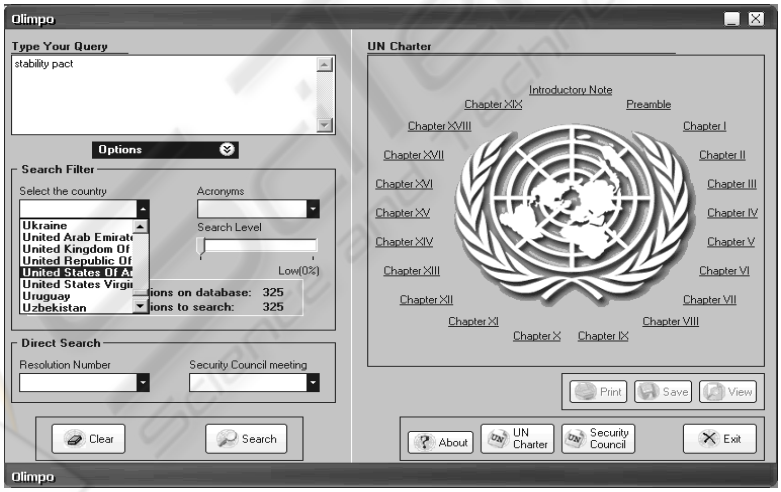

Figure 2: Olimpo Interface

In general, the system works in a way similar to other case-based systems [3] [5] (see Figure 3), where a manual entry passes through an adjustment and is then submitted to a comparison with the documents contained in the database, from which the most suitable ones are selected based on similarity calculations.

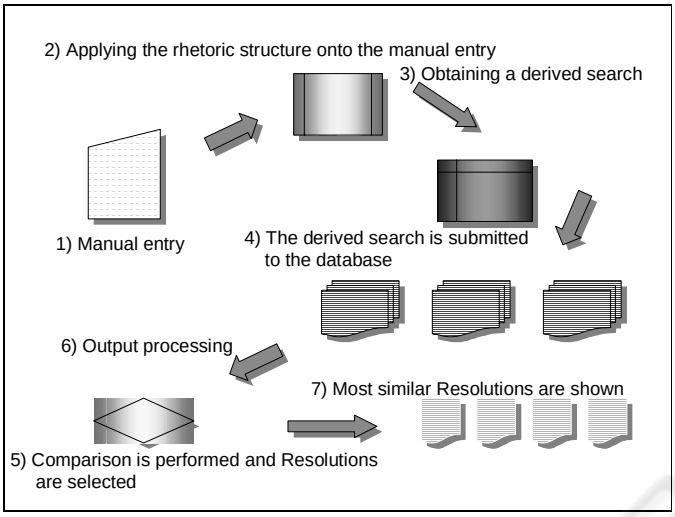

Figure 3: Searching Process

After a refined modelling of the database, the Resolutions are stored by Olimpo system, according to their characteristics and central attributes (main topics, related subjects, countries involved); peripheral attributes (other related Resolutions, other UNO´s organisms referred to); and superficial attributes (dates, numbers and names). This kind of structure allows to give (variable) weights to attributes, enabling a more precise, contextualised search.

Furthermore, the control of depth of search enables a selection of documents according to a higher or lower occurrence of indicative expressions within the text of the Resolution, before starting to compare the documents. This process provides a more efficient way of reducing the search field; it is not a mere pre-selection of documents based on their superficial characteristics, but a preliminary comparison oriented by the context related to the search input.

After completing the process, the result is a list of indicative expressions referring to the Resolutions, producing an individual record of the occurrence of each one of the expressions within the text of each Resolution. These records allow the system to make the comparison and to apply the global similarity metrics.

In addition to the indicative expressions, the process of automatic extraction of attributes was prepared to detect and extract the subject, date, number of the Resolution, acronyms, country names, and parts of the text that contain the expressions with higher occurrence.

The main features of the Olimpo system are the simultaneous use of textual information retrieving techniques based on CBR and the possibility of an extensive textual input. That makes the system to reach a differentiated performance in terms of information retrieval.

However, considering that the comparisons are based on a rhetoric structure previously prepared, the better working of the system is linked to a description of the search entry closer to that rhetoric structure. So, the system performance becomes gradually more consistent as the search entry language gets closer to the structure identified in the documents that generated the knowledge base of the system.

It has to be mentioned that all the Resolutions were monitored all the time with respect to the number of indicative expressions they presented during the structuring phase of the Resolutions knowledge base.

4 SIMILARITY METRICS

The similarity metrics was structured to consider the indicative expressions present in the case and in the search, after applying the rhetoric structure on the textual entry and producing the derived search. This derived search is actually the reference to work out the similarity metrics.

Taking as an example the case where a derived search with a total of 50 expressions is obtained after applying the rhetoric structure on a given search text: this set of expressions is compared to the records in the database and the similarity percentage is calculated based on the number of similar expressions found within each individual record. If 43 expressions are found, for instance, then the

similarity will be 86%; it will be 72%, if 36 expressions are found, or 56% in the case of finding just 28 expressions, and so on.

This type of metrics is quite simple, one of the most simple that could be used in this situation, but it works in a quite stable way and can be improved in the future by incorporating new mechanisms like trigrams or applying internal weights to the most frequent words found in the text of the Resolution.

In fact, what is the strong feature of the system is not the similarity metrics, but the way how the indicative expressions are organized so that the metrics provides a better performance.

A clear example of this particularity of the system is an expression formed by two words like "United Nations". A simple similarity based on counting individual words will show a 100% index when both words are found within the text, regardless their position, or 50% in the case just one of the words is found. However, if a differentiated indexation is used, by which "United Nations" (the two exact words appearing together) is one expression, "United" is another expression, and "Nations" is a third one (all with the same weight, for the time being), this configures a different situation. In this case, it is not enough finding the two words within the text, even when separated; they should appear together and having the exact meaning. Based on these criteria, the similarity index will be 33.33% when only one of the two words is found, or 66.66%, when both words are found in separate location, and it will reach 100% only when both words are present and appear together.

5 THE FUTURE OF OLIMPO SYSTEM

Internet was developed more than three decades ago, financed by the Department of the Defense of the U.S.A. Originally projected to connect the main systems of computers about a dozen of universities and organizations of research, the internet is currently accessible in hundreds of millions of computers in the world.

With the introduction of the World Wide Web - which allows personal computers to visualize documents based on multimedia about almost all the subjects - the internet literally explodes for what it certainly seems to become, the main mechanism of communication of the world.

Most computer applications were executed in computers which couldn´t communicate between themselves. Currently, applications that are communicated with hundreds of millions of computers of the world can be written. The internet establishes technology of computation and communications.

It turns our work easier. The information is accessible world-wide in an instantaneous and convenient form. It becomes possible that individuals and local small companies have a world-wide exposition. The people can find the better prices in products or services. Communities that have similar interests can stay in contact with each other. Researchers can be informed instantaneously about de last advances around the globe.

JAVA applications can be written in any computer platform, it means, any Java application needs a 32 bits version, such as Windows 95, Windows NT, other premium versions of Microsoft operational systems, MAC and UNIX, what results in a important economy of time and costs of development of systems for enterprises.

Java is a language completely object-oriented with hard support for proper techniques of engineering softwares. The programming object-oriented shapes objects of the real world with corresponding softwares. It takes advantage of the class relationships in which objects of a certain type - as a type of vehicles - they have the same characteristics. It also takes advantage of inheritance relationships where just-brought objects of a certain type inheritate characteristics of existing ranks, but still keeping exclusive characteristics. An object of the convertible rank certainly has the characteristics of the automobile class, but the ceiling of a convertible opens and closes.

The object-oriented programme (OOP) supplies us a more natural and intuitive way to see the process of programming – as follows, shaping objects of the real world, its attributes and its behaviors. The OOP also shapes the communication between objects. In the same ways as people change messages between themselves, the objects also communicate by messages.

The OOP encapsulates data (attributes) and methods (behaviour) in packages called objects, data and methods of an object are intimately united. The objects have the property of hiding the information. This means that even though the objects can communicate with each other through friendly interfaces, normally it is not allowed for objects to know how the other objects are implemented - the implementation details are occulted inside of the objects themselves. Certainly it is possible to direct a car without knowing well the details of how the engine, the transmission and the system of exhaust pipe work internally.

For the use on the WEB, the interfaces of Olimpo® system will be adapted in JAVA, because of the advantages shown above. The data base will

be extended, so as to enclose all the Resolutions of the Security Council, making it possible for the users of the Web to carry through consultations to the complete base and thus to take off more advantage of the system. The site will be of free access to the public, or either, without no cadastre form, and moreover it will possess a system of automatic updating to facilitate maintenance of information.

The system will have basically two distinct modules: the first one will be of the consultation that, based in the methodology of Context Structured Search – CSS will allow a similar retrieval to the existing one in the current system, containing filters and fields to open-search concept-based, and the administration module, who will be responsible for the inclusion of new resolutions and maintenance of the database. The administration module is necessary, because Olimpo® system has a base of knowledge especially developed to retrieve in an intelligent way the resolutions of Security Council. Knowledge base is a referential structure of representation in that domain specially studied, built so that the algorithms to retrieve information, can be semantic references in the search.

The knowledge base of Olimpo® system will be opened after its implementation to insert news data. Those can be given through the insertion of new resolutions, as well as more words and references in the specialized dictionary, which is meant for the mining and retrieval of information contained in the system database.

The updating of the data will be made in a dynamic form, that is, all the process will be made with a reduced effort, being up to the administrator of the system the function of including of new data in the base and the managing of the dynamics of the working of the system, keeping this up to date and compatible with the new referring events the Security Council of the ONU.

As follows below, there is a sketch of how the structure of the net will be . (Figure 4):

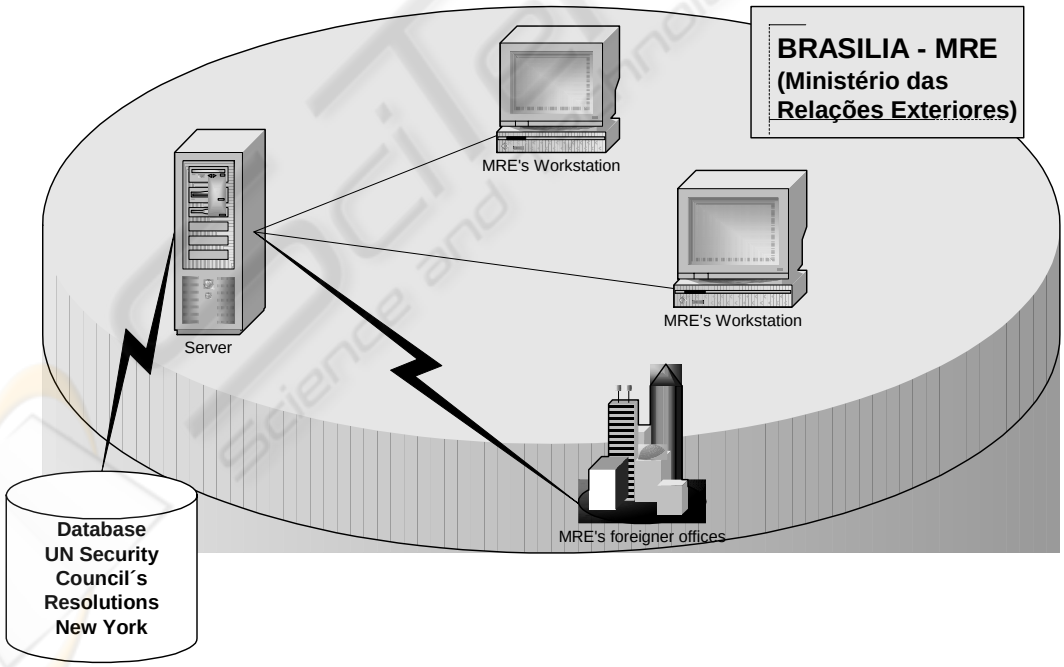

Figure 4

6 FORESEEN IMPACTS

Application of Olimpo® system will in such a way bring a series of impacts at national and international level, influencing not only in the technological field, but also in relation to citizenship, the knowledge and the research.

See some relevant impacts bellow:

Scientific Impact

- Accomplishment of seminaries and workshops on the applicability of the Technology of the Legal Information for the Security Council of the ONU;
- Publications in newspapers of the Institutions involved;
- Stimulus to the perfectioning of the techniques applied;
- Academic works about this theme.

Technological Impact

- Consolidation of the technique "Dynamically Contextualised Knowledge Representation" (RCKR), which increases performance of systems structured in knowledge;
- Perfectioning of the techniques of use of textual data bases with RCKD, that propitiate the application of PCE, for construction of Olimpo® system;
- Projection of Brazil as a technological reference.

Economic Impact

- Promotion of the companies of technology of the Information and Management of Knowledge;
- Dissemination of the applicability of the technologies used in Olimpo® by areas, that present the same difficulty in document research;
- Dissemination of the techniques of applied Management of the Knowledge to Olimpo®.

Social Impact

Facing the efficiency and the agility presented with the use of Olimpo® in the processes of search of information, increasing the search of the users to the available services.

7 CONCLUSION

Olimpo system is a clear example of an innovative approach to the issue of information retrieval from complex text databases. Based on CSS technology (Context Structured Search, the system reaches a higher performance using DCKR technique for knowledge representation.

The innovation and new features implemented represent an upgrade of Olimpo system, improving its overall performance and usability.

Olimpo® System will be an important Brazilian contribution for the unanimity between nations, in view of the fact that it democratizes the access to knowledge of the select circle of the Security Council of the ONU, what makes it an important technological contribution for the World-wide Peace. Faster, necessary and perfect decisions will have, surely, greater and better international legitimation, favoring the creation of a fairer scene. As a consequence, it will be an instrument for the demonstration of the technological capacity withheld by Brazil, in the area of softwares for the management of knowledge.

REFERENCES

Amondt, A.; Plaza, E. "Case-Based Reasoning: Fundamental Issues, Methodological Variations, and System Approaches". AI Communications 17(1), 1994.

Bench-Capon, T. J. M. Some observations on modelling case based reasoning with formal argument models. In: Proceedings of the Seventh International Conference on Artificial Intelligence and Law, p. 36-42, Oslo: Norway, June 14-18, 1999. 220 p.

Bruninghaus, Stefanie; ASHLEY, Kevin D. Toward adding knowledge to learning algorithms for indexing legal cases. In: Proceedings of the Seventh International Conference on Artificial Intelligence and Law, p. 9-17, Oslo: Norway, June 14-18, 1999. 220 p.

Bueno, Tânia Cristina D´Agostini. The use of juridical theory for retrieval from large juridical textual databases. Master Dissertation, PPGEP/UFSC. Florianópolis (Brazil): 1999. Original title: O uso da teoria jurídica para recuperação em amplas bases de textos jurídicos.

Bueno, Tania Cristina D'Agostini; Hoeschl, Hugo Cesar; Mattos, Eduardo da Silva; Barcia, Ricardo Miranda; Wangenheim, Christiane Gresse Von. JurisConsulto: Retrieval in Jurisprudencial Text Bases using Juridical Terminology. In: The Seventh International Conference on Artificial Intelligence And Law, 1999, Oslo. Proceedings of the Conference. New York: ACM, 1999. v.1. p.147-155.

Bueno, Tania Cristina D'Agostini; Hoeschl, Hugo Cesar; Mattos, Eduardo da Silva; Wangenheim, Christiane Gresse Von; Barcia, Ricardo Miranda. The use of juridical theory for retrieval from large juridical textual databases. In: Encontro Nacional de Inteligência Artificial, 1999, Rio de Janeiro. Anais do XIX Congresso Nacional da Sociedade Brasileira de

Computação. Rio de Janeiro: Edições EntreLugar, 1999. v.4. p.107-120. Original title: Uso da teoria jurídica para recuperação em amplas bases de textos jurídicos.

Bueno, Tania Cristina D'Agostini; Hoeschl, Hugo Cesar; Mattos, Eduardo da Silva; Barcia, Ricardo Miranda; Bortolon, André; Wangenheim, Christiane Gresse Von. JurisConsulto. Florianópolis (Brazil): 1999. Software rights registered.

Hoeschl, Hugo Cesar. Olimpo System: Juridical Information Technology for UNO´s Security Council. Florianópolis (Brazil): UFSC, 2002. Doctorate Thesis. Original title: Sistema Olimpo: Tecnologia da Informação Jurídica para o Conselho de Segurança da ONU.

Hoeschl, Hugo Cesar; Barcia, Ricardo Miranda; Bueno, Tânia Cristina D´Agostini; Mattos, Eduardo da Silva; Bortolon, Andre; Donatti, Fabrício Tadeu. Olimpo System. Florianópolis (Brazil), 2000. Software rights registered.

Hoeschl, Hugo Cesar; BUENO, Tania Cristina D'agostini; MATTOS, Eduardo da Silva; BORTOLON, André; RIBEIRO, Marcelo Stopanowski; THEISS, Irineu; BARCIA, Ricardo Miranda. STRUCTURED CONTEXTUAL SEARCH FOR THE UN SECURITY COUNCIL. In: ICEIS - 5TH INTERNATIONAL CONFERENCE ON ENTERPRISE INFORMATION SYSTEMS, 2003, Angers. Proceedings of the fifth International Conference On Enterprise Information Systems. Setúbal: School of Technology of Setúbal, 2003. v. 2, p. 100-107. Referências adicionais: Classificação do evento: Internacional; França/Inglês; Meio de divulgação: Impresso; Homepage: http://www.iceis.org/papers.htm; ISSN/ISBN: 9729881618.

Hoeschl, Hugo Cesar; BUENO, Tania Cristina D'agostini; BORTOLON, André; RIBEIRO, Marcelo Stopanowski; MATTOS, Eduardo da Silva; THEISS, Irineu. Dynamically contextualized knowledge representation of the United Nations Security Council Resolutions. In: NINTH INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW, 2003, Edimburgo. ICAIL 2003 Proceedings. New York: ACM, 2003. v. 1, p. 95-96. Referências adicionais: Classificação do evento: Internacional; Escócia/Inglês; Meio de divulgação: Impresso; Homepage: http://www.cirfid.unibo.it/~agsw/icail03/; ISSN/ISBN: 1581137478.

Hoeschl, Hugo Cesar; Bueno, Tânia Cristina D´Agostini; Mattos, Eduardo da Silva; Bortolon, Andre; Barcia, Ricardo Miranda. Olimpo: Contextual Structured Search to improve the representation of UN Security Council Resolutions with information extraction methods. In: The 8th International Conference on Artificial Intelligence and Law, 2001, St.Louis, MO, USA. Proceedings of the Conference. New York: ACM, 2001. v.1. p. 271-218.

Steinfus, Ricardo. Handbook of International Organizations. Porto Alegre (Brazil), 1997, 352p. Original title: Manual de organizações internacionais.