AN INTELLIGENT SEARCH ENGINE FOR ELECTRONIC GOVERNMENT APPLICATIONS FOR THE RESOLUTIONS OF THE UNITED NATIONS SECURITY COUNCIL

Hugo C. Hoeschl, Tânia Cristina D. Bueno, Andre Bortolon, Eduardo S. Mattos, Marcelo S. Ribeiro, Irineu Theiss and Ricardo Miranda Barcia
E-Gov, Juridical Intelligence and Systems Institute – Ijuris, Florianópolis, Santa Catarina - Brasil, tania@ijuris.org, digesto@digesto.net, bortolon@wbsa.com.br, mattos@wbsa.com.br; marcelo@wbsa.com.br, irineu@wbsa.com.br, rbarcia@eps.ufsc.br, http://www.ijuris.org

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. It is included an explanation about the UN Security Council itself and how it works, as well as a detailed analysis of the format of the resolutions and its main characteristics. 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).

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

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 items 2 and 3 of this paper address the UN Security Council and the Resolution document; in item 4 the knowledge representation methodology is presented; in items 5 and 6 the Olimpo system is described and its performance is analysed; in items 7 to 9 characteristics and new features of the system are described; and item 10 is the conclusion of the paper.

2.   ABOUT THE UN SECURITY COUNCIL

As a consequence of the current world context, the UN Security Council is at the highest evidence. The international media is proving that. Taking as an example the date 18th of September 2002, the Security Council was on the front page of some of the most important newspapers worldwide:

- The Washington Post: "U.N. Questions Need for New Resolution on Iraq";
- Le Monde: "Irak: Division au sein du Conseil de sécurité";
- Independent: "UN split over Iraqi offer";
- El Pais: "Bush desprecia la oferta de Irak e insiste en que la ONU debe actuar";
- The Times: "Saddam offer tests fragile alliance";
- Clarín: "Bush ignora la oferta de Irak: 'Es hora de actuar'".

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, as shown on Figure 1. 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.
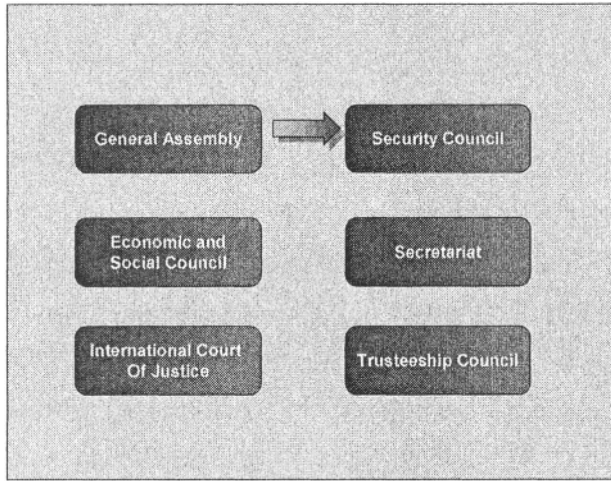
Figure 1. UNO main bodies issuing resolutions (Source: www.un.org/documents/index.html)

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 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 bod-

ies with specific power delegation to perform certain tasks, on a permanent or "ad hoc" level, like the sanctions committee, as shown on Figure 2.
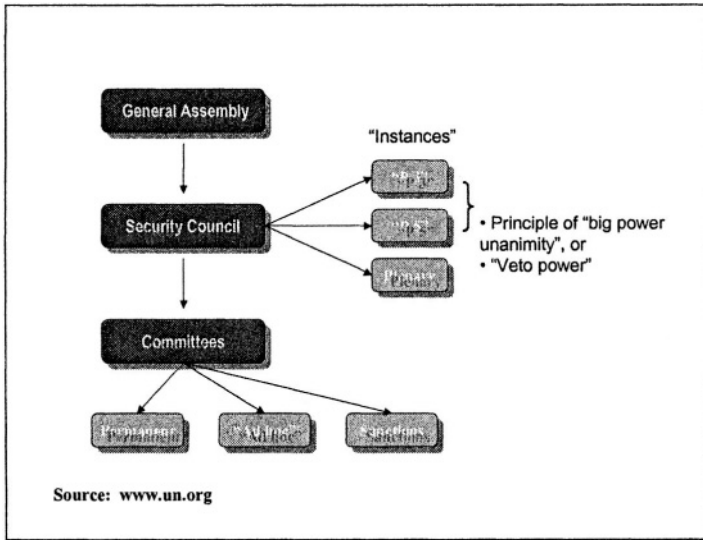

Figure 2. Position of the Security Council

Among the documents issued by the Security Council, six of them have greater relevance, as shown on Figure 3. 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.
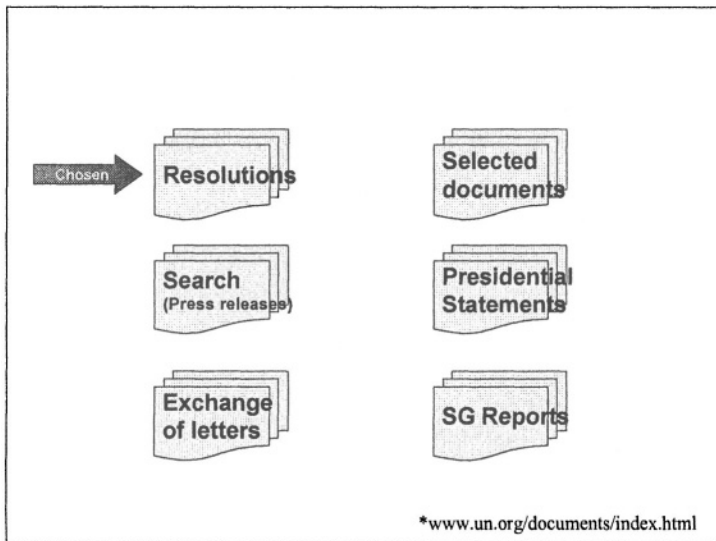
Figure 3. Types of documents issued by the Security Council

3. ABOUT THE RESOLUTIONS

Many different aspects surround the current debate on the Iraqi crisis, but one is definitely important: the document that will support the decision on the issue will be a Resolution of the Security Council. It is necessary to learn how these documents look like and which is their writing format.

As shown on Figure 4, the official UNO's site shows a specific section about documents designated as Documentation Centre, which maintains up-dated documents of UNO's bodies, including the Security Council.
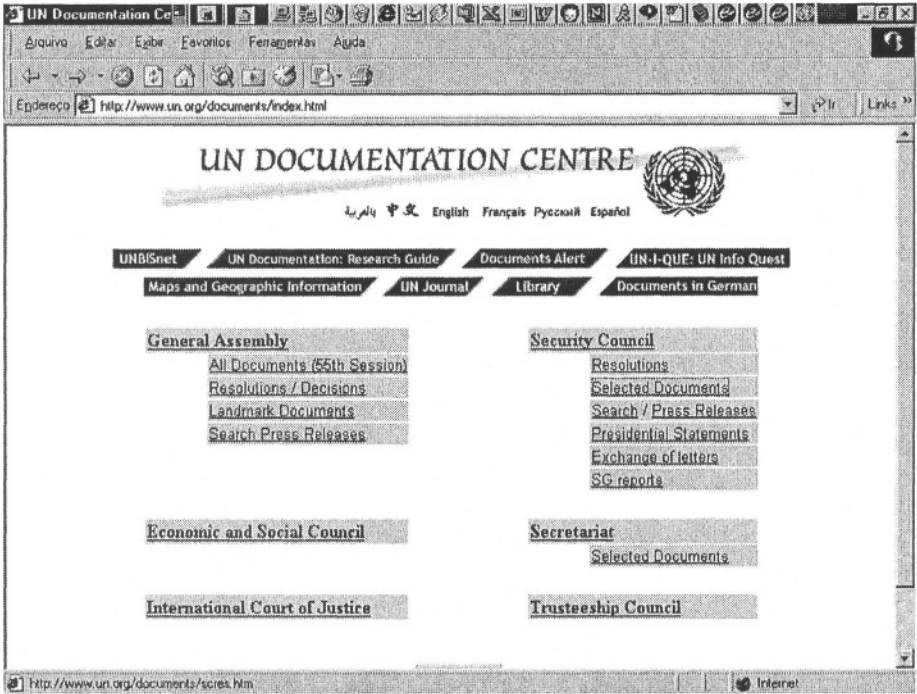
Figure 4. UN Documentation Centre

It was interesting to notice that the Security Council had the largest variety of documents among the different UN bodies and this confirmed the relevance of structuring the information belonging to that body.

Once the source of documents was selected, it started the process of capturing the documents on the web to build the knowledge base of the system.

To get a better understanding of the structure of a Resolution, it is useful to take a look at one of these documents (or part of it). It is shown below part of the text of Resolution number 1244, issued in 1999. This Resolution was chosen because it presented a high number of indicative expressions (a total of 137 expressions).
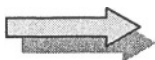
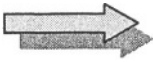**"Security Council resolution 1244 (1999) on situation relating to Kosovo**

United Nations
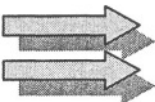
S/RES/1244 (1999)

10 June 1999

RESOLUTION 1244 (1999)

Adopted by the Security Council at its **4011th meeting,**
on 10 June 1999

The Security Council,
Bearing in mind the purposes and principles of the Charter of the United Nations, and the primary responsibility of the Security Council for the maintenance of **international peace** and security,
Recalling its resolutions 1160 (1998) of 31 March 1998, 1199 (1998) of 23 September 1998, 1203 (1998) of 24 October 1998 and 1239 (1999) of 14 May 1999,
Regretting that there has not been full compliance with the requirements of these resolutions,
Determined to resolve the grave **humanitarian situation** in Kosovo, **Federal Republic of Yugoslavia,** and to provide for the safe and free **return of all refugees** and **displaced persons** to their homes,
Condemning all **acts of violence** against the Kosovo population as well as all **terrorist acts** by any party,
Recalling the statement made by the **Secretary-General** on 9 April 1999, expressing concern at the humanitarian tragedy taking place in Kosovo,
Reaffirming the right of all refugees and displaced persons to return to their homes in safety,

(..........)

10. Suspension of **military** activity will require acceptance of the principles set forth above in addition to **agreement** to other, previously identified, required elements, which are specified in the footnote below.1 A **military-technical agreement** will then be rapidly concluded that would, among other things, specify additional modalities, including the roles and functions of Yugoslav/Serb personnel in Kosovo:
Withdrawal
- Procedures for withdrawals, including the phased, detailed schedule and delineation of a buffer area in Serbia beyond which forces will be withdrawn;
Returning personnel
- Equipment associated with returning personnel;
- Terms of reference for their functional responsibilities;
- Timetable for their return;
- Delineation of their geographical areas of operation;
- Rules governing their relationship to the international security presence and the **international civil** mission.

The arrows indicate, respectively, the subject, number of the Resolution, issue date, and the beginning of the text from which the indicative expressions (in bold) were extracted.

4.   DYNAMICALLY CONTEXTUALISED KNOWLEDGE REPRESENTATION

Olimpo's performance is centred around the combination of aspects derived from CBR and text information retrieval, in addition to an adequate organisation 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 organisation 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.

In general, the system works in a way similar to other case-based systems [3] [5] (see Figure 5), 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.

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.
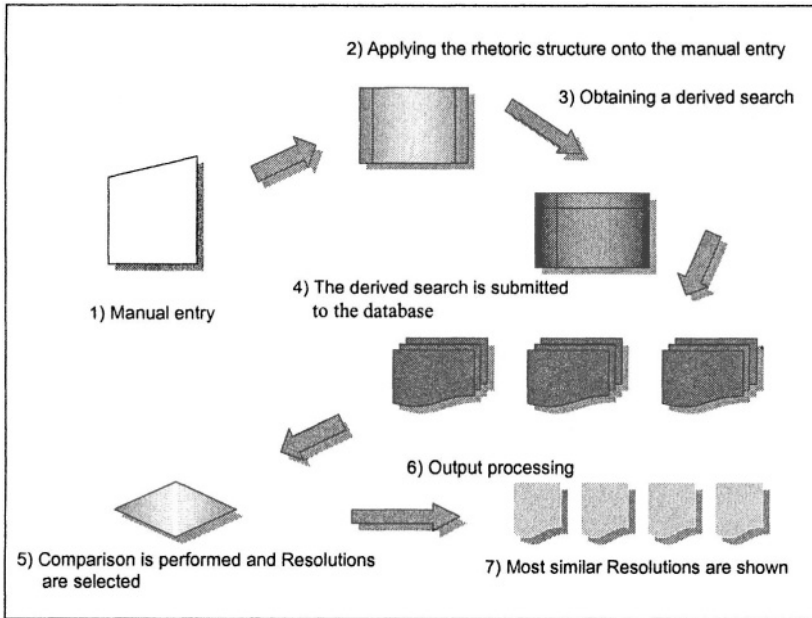
Figure 5. Searching Process

5.   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 by Figure 6 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.
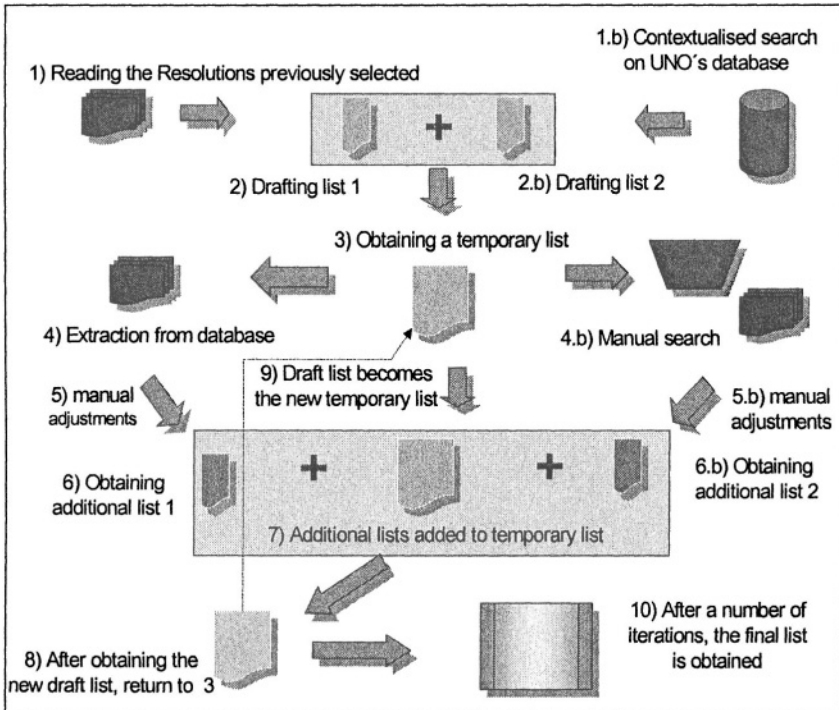

Figure 6. Analysis of the Rhetoric Structure

6.   CHARACTERISTICS OF THE DOCUMENTS

As to the structure of the document, UNO's Security Council Resolutions have some characteristics that facilitate the application of the technology presented in this paper. They are the following:

- Text written in homogeneous format – standard format used for all the documents;
- Standardised rhetoric structure – the domain vocabulary is quite specific and restrict, defined by Organisation's attributions as stated in the United Nations Charter;

- The language of the domain needs to be simplified to turn it more usual – the Resolutions are public, official documents (they need to have an usual language so that information can be retrieved when accessed by people).

7.   STRUCTURED CONTEXTUAL SEARCH – SCS

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

- • For building the rhetoric structure of the system, it is taken into consideration the context of the stored documents;
- • This context is the basis for the input adjustment process, as well as for the comparison and selection of documents;
- • 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.

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.

8.   OPERATION AND PERFORMANCE OF THE OLIMPO SYSTEM

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.

9.   THE OLIMPO SYSTEM

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 Figure 7).
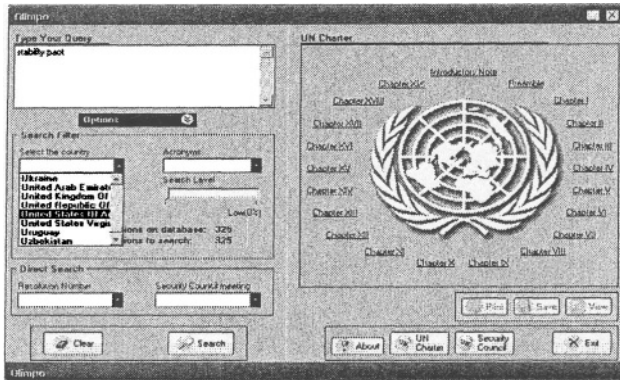

Figure 7. Olimpo Interface

10.   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 expressionsis 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.

11.   COMPOSING THE RECORDS

Besides searching for indicative expressions, the process of automatic extraction of attributes was prepared to detect and extract the subject, date, number of the Resolution, acronyms, name of countries, number of decisions, as well as parts of the text containing the expressions with higher ocurrence within the text, as shown on Table 1

Table 1. Automatic extraction of attributes

Index	Description	Importance of retrieval
Subject	Short description about the most important situation discussed in the Resolution	Used to know the Resolution subject
Date	Year of issuance of the Resolution	Determine a specific year of interest
Resolution Number	States the number supplied by de UN Security Council	Used to identify the Resolution
Meeting	States de number of the meeting when the Resolution was issued	Can be used to retrieve the Resolution by refering to the meeting
Countries	States the countries involved in the issue	Used to know the countries involved in the conflict/subject
Decisions	Number of different decisions in the Resolution	Can be used to understand the extension and complexity of the Resolution
Acronyms	Acronyms of official and unofficial institutions	Can be used to retrieve the institutions related to the Resolution. E. g.: OTAN
Indicative Expressions	Determines the most relevant information in the Resolution and its re-definitions	Used to retrieve the Resolutions based on similar topics

The analysis taking into consideration all these information enables the documents to be compared, in a precise way, among themselves, with other texts, or with the text of the query entered by the user.

12.   COMPARATIVE ANALYSIS: DATABASE SEARCH TOOL VERSUS OLIMPO

When a random input is typed, the chance of an effective retrieval is lower than in the case of an input based on a text written in a language coincident with that already identified, for example, the text of a Security Council's Resolution. For random inputs, the conventional database presents a superior performance; but this changes as the search text becomes closer to the language used in the documents contained in the system database.

The possibility of having an input with higher volume of text (2,300 words, 15,000 characters or 270 lines of text) becomes an important differential factor for the Olimpo system. The output obtained with the traditional database tool is quite limited when compared with the output produced by Olimpo.

As illustrated on Figure 8, Olimpo's efficiency increases as the number of words in the input text increases. On the other hand, the efficiency of the database search tool decreases sharply as the volume of input text increases.
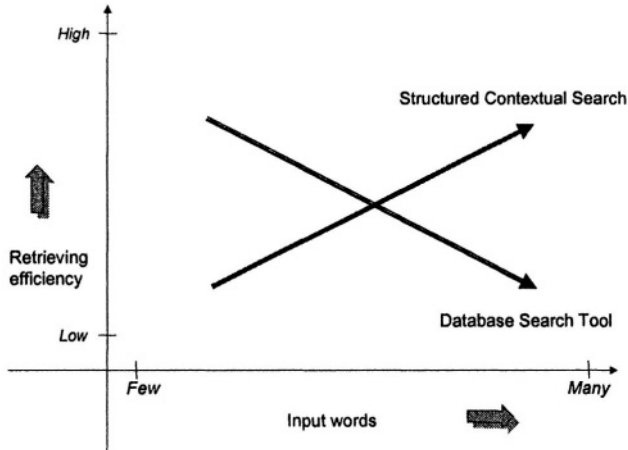

Figure 8. Olimpo System Retrieving Efficiency

13.   SEARCH LEVEL

An important feature of the Olimpo system is the possibility of previously filtering the documents to be included in the similarity metrics.

The filtering process, designated as search level, is described by Figure 9. It reduces the amount of records to be searched, but it is not a simple partition of the database. It is not a question of format, but is based on "merit" issues, enabling a preliminar look at the searching universe even before starting the search.
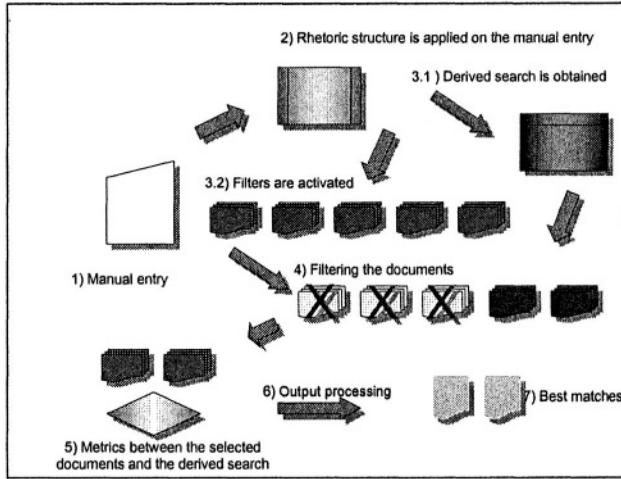
Figure 9. Searching process (with filters and pre-search)

By controlling the depth of the search level it is possible to select documents based on the number of indicative expressions found within each document before starting to compare them. This process is described by Figure 10 and it provides a reduction of the searching universe in a more efficient way. It is not about simply selecting documents according to superficial characteristics, but it makes possible to have a preliminary comparison oriented by the context related to the query entered by the user. This process enables a significant improvement of the system performance.
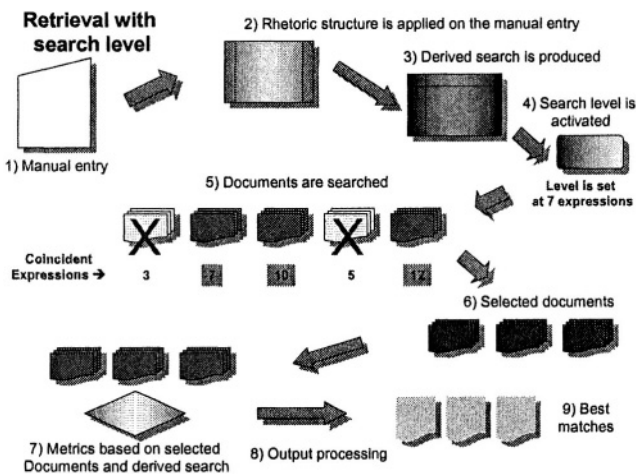

Figure 10. Retrieval with search level

14.   SYSTEM STATISTICS

The statistical analysis of the relationship between concepts and Resolutions provides relevant conclusions. The average of indicative expressions per Resolution is 55.18, which is an impressive number when considering that this average would never be over 10 using the former methodology. That average should increase significantly after completing the development of the system and including all the Resolutions in the system database.

Among the Resolutions with the best performance, 12 of them were selected showing over 100 indicative expressions found (Table 2).

Table 2. Occurrence of indicative expressions

Order	Resolution number	Ocurrences
1	1244	137
2	1075	124
3	1265	123
4	1234	120
5	1087	119
6	1247	113
7	1199	112
8	1124	111
9	1174	111
10	1072	106
11	1284	106
12	1203	100

The usual language of the Security Council includes the frequent ocurrence of acronyms. The different number of acronyms occuring within the text of a Resolution is counted; Table 3 shows the 10 Resolutions with the highest number of acronyms.

The meaning of this information is linked to the institutional context of a Resolution. Probably the one making reference to a higher number of acronyms (among those commonly used by the Security Council) involves a subject with stronger relation to different institutions.

Table 3. Occurrence of acronyms

Resolution	Acronyms
1075	6
1045	5
1063	5
1087	5
1118	5

1035	4
1048	4
1124	4
1127	4
1195	4

15.   NEW FEATURES OF THE OLIMPO SYSTEM

The Olimpo system was the subject of a paper presented at ICAIL 2001, in St. Louis [10]. Since then significant changes have been implemented as described above. Besides them a number of minor features have been implemented, turning the system more complete, more user-friendly and more dynamic. These new features help to improve the general performance of the system and include the following:

- • A new interface allowing the user to visualize, at the same time, the search query and the result produced by the system;
- • The use of text mining to generate statistics from texts shown in a graphical form;
- • More Resolutions were added to the system database, including the whole set issued in 2002 and the most recent ones issued in 2003;
- • The United Nations Charter was added as a document available for direct searching;
- • The Provisional Rules of Procedure of the Security Council were also added to the document base for direct searching purposes.

16.   CONCLUSION

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

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

REFERENCES

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

2. 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.
3. 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.
4. 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.
5. 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.
6. 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.
7. Bueno, Tania Cristina D'Agostini; Hoeschl, Hugo Cesar; Mattos, Eduardo da Silva; Barcia, Ricardo Miranda; Bortolon, André; Wangenheim, Christiane Gresse Von. Juris-Consulto. Florianópolis (Brazil): 1999. Software rights registered.
8. 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.
9. 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.
10. 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.
11. STEINFUS, Ricardo. Handbook of International Organizations. Porto Alegre (Brazil), 1997, 352p. Original title: Manual de organizações internacionais.