A Rough Sets Approach for Relevant Internet/Web Online Searching

Erika Martinez Ramirez and Rene V. Mayorga

***Abstract*—**The internet is constantly expanding. Identifying web links of interest from web browsers requires users to visit each of the links listed, individually until a satisfactory link is found, therefore those users need to evaluate a considerable amount of links before finding their link of interest; this can be tedious and even unproductive. By incorporating web assistance, web users could be benefited from reduced time searching on relevant websites. In this paper, a rough set approach is presented, which facilitates classification of unlimited available e-vocabulary, to assist web users in reducing search times looking for relevant web sites. This approach includes two methods for identifying relevance data on web links based on the priority and percentage of relevance. As a result of these methods, a list of web sites is generated in priority sequence with an emphasis of the search criteria.

***Keywords*—**Web search, Web Mining, Rough Sets, Web Intelligence, Intelligent Portals, Relevance.

I. Introduction

ONLINE Online searches require a considerable amount of time from the user. In addition, every minute, countless numbers of new e-sites are added to the web. This results in an increase of e-vocabulary and ambiguousness to the meaning of search terms from the internet. Past research has focused on solving the problem of classifying web documents using the *similarity of terms* (vector space model)[5][7]; *keyword mapping* of terms[6][18]; and classification through *summarization* [17] among other methods provided by the literature. Some researchers consider the *term frequency* (TF) presented in those sites for the classification process [2]. On the other hand, large amounts of irrelevant repetitive terms are found in some websites, which are 'hidden' in the content of the site. For example, terms may be hidden in the background of the website, using the same text colour as the background [3]. The 'hidden' terms limit web classification based on TF [3], considering that an option for *web browsers* to rank websites is according to their terms and their frequency. Other researchers have considered the brief description (*snippet*), provided by the browser, as the "document". This "document" is used to obtain terms and determine the frequency of terms within the website.

The goal of this research is to facilitate the process of searching for relevant online information and to reduce the time required identifying relevant links. In order to achieve this goal, two methods are presented to classify in priority sequence based on a Rough Set approach. In order to identify the most relevant websites to a specific set of terms or queries, a new approach is proposed based on Rough Sets (RS). This approach aims to define the conditional attributes and decision attributes from the web "document", identify the relevance of web links for a query, and classify the links based on their relevance.

The query terms are enumerated from *Query Term 1* to the total number of terms *n*, which are specified with the following notation $Qt_1 \quad \cdots \quad Qt_n$. These query terms, named *search query*, are sent to the *search engine* and return as a set of links ordered by the browser. This is illustrated in Fig 1.

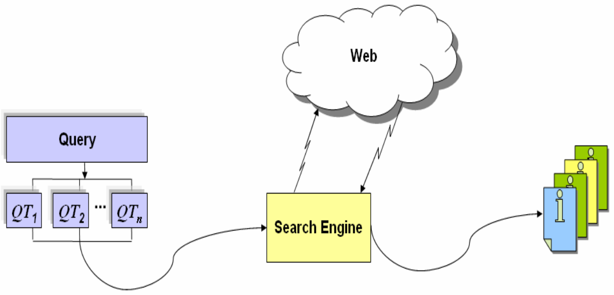


Fig. 1 Define *search query* and perform search. The set of Query Terms (QT_1, QT_2, … QT_n) are sent to the Search Engine, which obtain the title, URL and snippet of the web links

For each of the websites obtained, the *search engine* provides the *title* of the site, the *URL* address and the "document" (snippet) briefly describes the site. The sites obtained are ranked by the *web browser* and may include redundant data or repetitive URLs.

II. Artificial Intelligent Technique: Rough Sets

Several Artificial Intelligent (AI) techniques have been studied toward select an appropriate technique to allow classify relevant web links. Techniques previously studied and/or used in our research [12] has shown us that Fuzzy

Dr. Erika Martinez Ramirez is in the Information Technology Department at Agriculture Financial Service Corporation, Lacombe, Alberta, Canada.

Dr. Rene V. Mayorga heads the Wise & Intelligent Systems & Entities Laboratory in the Faculty of Engineering (Industrial Systems Engineering) at the University of Regina, SK, S4S 0A2; Canada. Phone: (306) 585-4726; Fax: (306) 585-4855; e-mail: Rene.Mayorga@uregina.ca).

Logic processes imprecise data through Membership Functions (MFs) [8]; Expert Systems are based on sets of rules [1]; Neural Nets are based on train a net according to inputs and outputs [8]; and Hybrid Systems combine two or more techniques in their process. Consequently, other techniques were explored for data classification including: Rough Sets and Granular Computing. Rough Sets (RS) is a useful tool to classify features or data. RS [9][14]. Moreover, RS has the ability to simplify information systems through computational reduction of conditional attributes, eliminate duplicate rows, and eliminate superfluous values of attributes [14]. RS facilitates the process of reducing large amounts of ambiguous data often provided by web browsers.

A. *Rough Sets*

The selection of relevant data and reduction of time consumption is based on the analysis of the original rough set theory [14]. RS begins defining the Information System (IS), which is the main factor in identifying the data from the website. The information system is defined as the decision table that contains conditions and decisions attributes, represented by the Equation 1.

$$\mathrm{A} = (U, A \cup \{d\}) \qquad (1)$$

where U is the Universe, *A* is the conditional attributes and *d* is the decision attributes. In our research work, attributes are obtained from the frequency of terms which occurs in snippets. The following example illustrates an Information System indicating the availability of products.

The indiscernibility relation is represented by the equation 2 and using an example in equation (3)

$$[x]_{IND(P)} = \bigcap_{R \in P} [x]_R \qquad (2)$$

and described as follow:

Red → x_1, x_3, x_7 Round → x_1, x_5
Blue → x_2, x_4 Square → x_2, x_6, x_8
Yellow → x_5, x_6, x_8 Triangular → x_3, x_4, x_7

Small → x_1, x_4, x_5, x_6, x_8
Large → x_2, x_3, x_7

$$\begin{aligned} U / R1(color) &= \{\{x_1, x_3, x_7\}\{x_2, x_4\}\{x_5, x_6, x_8\}\} \\ U / R2(shape) &= \{\{x_1, x_5\}\{x_2, x_6, x_8\}\{x_3, x_4, x_7\}\} \\ U / R3(size) &= \{\{x_2, x_3, x_7\}\{x_1, x_4, x_5, x_6, x_8\}\} \end{aligned} \qquad (3)$$

The indiscernibility partition is represented as in Table I.

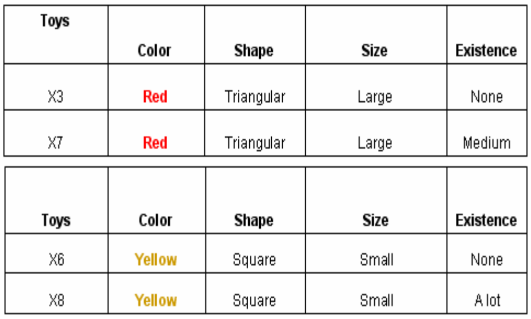
The approximation of sets is made by defining the lower and upper boundaries [9] as follow:

Lower $\underline{R}X = \cup\{Y \in U / R : Y \subseteq X\}$

Upper $\overline{R}X = \cup\{Y \in U / R : Y \cap X \neq 0\}$

R-boundary $\overline{R}X - \underline{R}X$

TABLE I
INDISCERNIBILITY PARTITION



Toys	Color	Shape	Size	Existence
X3	Red	Triangular	Large	None
X7	Red	Triangular	Large	Medium

Toys	Color	Shape	Size	Existence
X6	Yellow	Square	Small	None
X8	Yellow	Square	Small	A lot

The reduction of *knowledge* is the minimal set of attributes obtained without compromising the classification and consistency of the information system, based on the reducts and core [14].

If Q is a REDUCT of P and R $\in$ P-Q, then *IND*(P) = *IND*(Q), Q $\subseteq$ P - {R} $\subseteq$ P

If Q is a CORE of P is $CORE(P) = \cap RED(P)$

For simplification of Decision Tables (Information Systems), the following steps are required: computation of reducts of conditional attributes, elimination of duplicate rows, and elimination of superfluous values of attributes.

B. *Definition of Conditional Attributes and Decision Attributes*

The matrix of conditional attributes is created based upon the terms contained in each “document” and the number of times they appear in each websites. This is summarized in a matrix of term frequencies of [*l* x *c*] conditional attributes, as follows in equation (4):

$$\begin{array}{cccc} T_1 & T_2 & \cdots & T_c \\ \left[\begin{matrix} f_{11} \\ f_{21} \\ \vdots \\ f_{l1} \end{matrix}\right. & \begin{matrix} f_{12} \\ f_{22} \\ \vdots \\ f_{l2} \end{matrix} & \begin{matrix} \cdots \\ \cdots \\ \ddots \\ \cdots \end{matrix} & \left.\begin{matrix} f_{1c} \\ f_{2c} \\ \vdots \\ f_{lc} \end{matrix}\right] \end{array} \qquad (4)$$

Where

$T_1 \cdots T_c$: terms contained in the “document” websites obtained.

$f_{11} \cdots f_{1c}$: term frequencies in “document”.

c : number of conditional attributes, terms.

l : number of cases, links.

The decision attributes for the information system is summarized in a matrix [*l* x *d*], represented as follows in equation (5):

$$\begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1d} \\ d_{21} & d_{22} & \cdots & d_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ d_{l1} & d_{l2} & \cdots & d_{ld} \end{bmatrix} \quad (5)$$

Where

$d_{11} \cdots d_{1d}$	:	decision attributes for each website.
d	:	number of decision attributes.
l	:	number of cases, links.

III. Relevant Identification Approach

The creation of the decision attributes is the primary part of the Decision Process of Relevance. For this reason, the creation and explanation of those attributes are described in this section.

A. *Decision Process of Relevance*

Current research focuses upon the classification of web links using the vector space model [2] to find similarities between terms; using tolerance rough set [20] and tolerance rough set clustering models [13] to model relationships between terms and documents; using clusters/clustering [20]; Fuzzy Rough Sets [4] [15]; and Rough Set Clustering [10] in order to classify websites according to their terms. While classifications are performed to categorize related websites, our research classifies the websites in categories according to their relevance (research not shown).

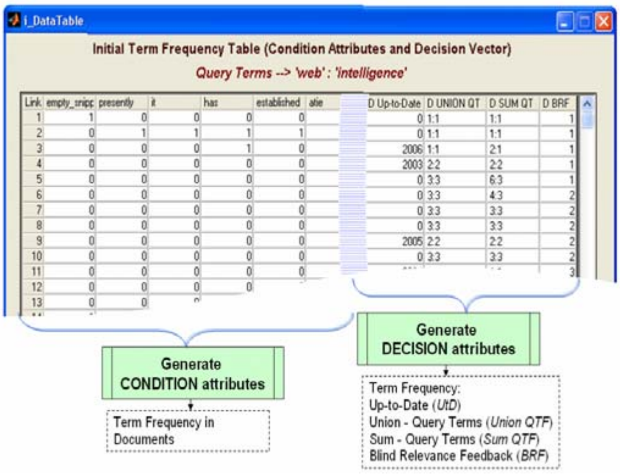
Furthermore, some literature addresses relevance of web data retrieval [2] and email classification [13], which requires user interaction, which is called User Relevance Feedback (URF). Their systems proposed a set of suggested terms and/or clusters. Users give their feedback as to whether the clusters/terms were relevant. The accuracy of the method for identifying relevance has been strongly related to the users' knowledge of the topic [18]; therefore, user feedback may not always give a correct degree of relevance. Due to this fact, some researchers incorporate the use of Blind Relevance Feedback (BRF) to indicate the relevance of the websites [6]. BRF indicates relevance based on the rank given to the site by the *web browser*. Some researchers focused on the way to address relevance on web documents without considering user interaction [16] [19]. This research addresses this relevance through a proposed Decision Process of Relevance and retains to the user interaction until completing the process.

In order to define the decision attributes, more decision parameters must be considered. These parameters assist the search for relevant online information using RS. Such attributes are *Up-to-Date* (*UtD*) web links, Occurrence of the *Union of Query Terms Frequency* (*Union QTF*), Occurrence of the *Sum of Query Terms Frequency* (*Sum QTF*), and the *BRF*.

- The *Up-to-Date* attribute of the web links is indicated by the time indices found in the web "document".
- The *Union QTF* attribute considers the number of occurrences where all the terms contained in the *search query* appear in the "document". This enhances the relationship between the "document" and the *search query*, assuring all query terms are uniformly associated with each web "document".
- The *Sum QTF* attribute involves all occurrences of either query term in the "document". This attribute represents the relationship between the "document" and the query terms.
- The *BRF* attribute defines the relevance feedback based on the given rank to the site assigned by the *web browser*. As a consequence, the first links listed will have a higher rank of BRF while the last links will have a lower rank. Websites with the highest BRF score receive the value of one and this value increases according to the number of websites, meaning that the level of BRF is lower, as follows:

$$\begin{matrix} 1 & \ldots & \text{Highest BRF} \\ \vdots & \ddots & \vdots \\ n & \ldots & \text{Lowest BRF} \end{matrix}$$

Having defined the conditional and decision attributes, the Initial Term Frequency Table is created. This table contains all the terms found in the "documents" and their frequencies (Condition Attributes), and the correspondent Decision Attributes for each web link. Fig. 2 contains these attributes.



Initial Term Frequency Table (Condition Attributes and Decision Vector)

Query Terms --> 'web' : 'intelligence'

Link	empty_snipp	presently	it	has	established	atie	D Up-to-Date	D UNION QT	D SUM QT	D BRF
1	1	0	0	0	0		0	1:1	1:1	1
2	0	1	1	1	1		0	1:1	1:1	1
3	0	0	0	1	0		2006	1:1	2:1	1
4	0	0	0	0	0		2003	2:2	2:2	1
5	0	0	0	0	0		0	3:3	6:3	1
6	0	0	0	0	0		0	3:3	4:3	2
7	0	0	0	0	0		0	3:3	3:3	2
8	0	0	0	0	0		0	3:3	3:3	2
9	0	0	0	0	0		2005	2:2	2:2	2
10	0	0	0	0	0		0	3:3	3:3	2
11	0	0	0	0	0					3
12	0	0	0	0						
13	0	0								

Fig. 2 Table with Conditional and Decision attributes.

B. *Algorithm to identify relevant web data*

In order to begin the decision process for relevance, a set of query terms is required. In this implementation, two options are available to specify a query: (1) the user indicates a set of terms to be searched for or (2) the set of terms are loaded from a list of predefined categories. This approach consists of classifying links in order to obtain an ordered and reduced set of the most relevant web links, as illustrated in Fig. 3.

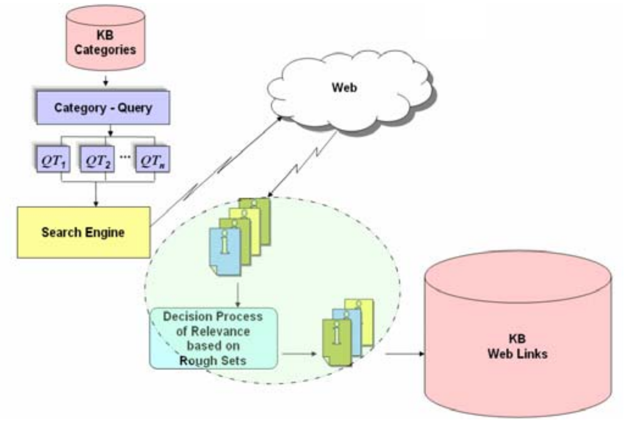


Fig. 3 General algorithm to classify web links

The rough set approach for reducing web search times is exemplified in Fig. 4, where the inputs are the list of websites ranked from by browser and the result are a classified list of relevant websites.

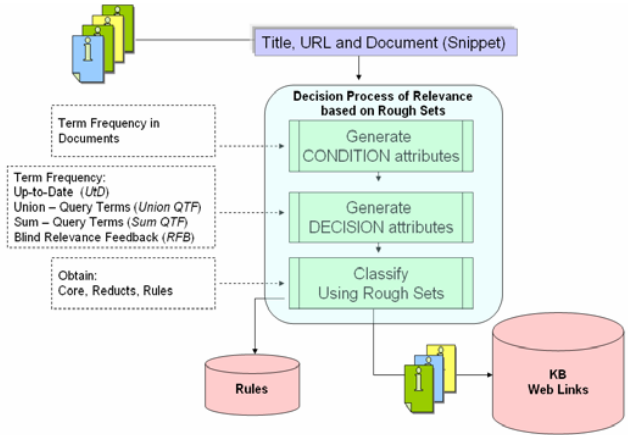

Title, URL and Document (Snippet)
Decision Process of Relevance based on Rough Sets
Generate CONDITION attributes
Generate DECISION attributes
Classify Using Rough Sets
Term Frequency in Documents
Term Frequency:
Up-to-Date (UtD)
Union – Query Terms (Union QTF)
Sum – Query Terms (Sum QTF)
Blind Relevance Feedback (RFB)
Obtain:
Core, Reducts, Rules
Rules
KB Web Links

Fig. 4 Rough set approach for reducing web search times

Furthermore, to this approach, the interaction and feedback from the user is required once the decision process of relevance has been completed. Once the decision attributes have been defined, and the decision processes for relevance accomplished, two methods are proposed to assign the degree of relevance to the websites, these are:

Priority of Decision Attributes (PDA). In this method, each of the decision parameters receives a level of priority, i.e. *Up-to-Date* attribute of the web links receives the first priority, followed by the *Union QTF* attribute and then the *Sum QTF* attribute of term frequencies and, lastly, the *BRF* attribute. Therefore, the classification of websites is performed considering the level of priority previously mentioned.

Percentage of Relevance of Decision Attributes (PRDA). This method, in opposition to the other method, proposes defining the level of priority by specifying the percentage of relevance for each decision attribute. This percentage varies according to the preference of information which is searching for. These two methods are described in more detail in the Evaluation of the System Performance.

The terms “Web Intelligence” is the *search query* selected as an example to demonstrate the effectiveness of our proposed research. The obtained results from this query generate the corresponding tables for these two methods and are described and analyzed in the following section.

IV. Implementation

The developed research was implemented in a Matlab environment; however other software development environments could be used. Matlab framework allows to create and personalize commands and program code and to create independent and customized functions through the M-file editor. The Rough Sets (RS) algorithms and methods of relevance were programmed on this editor. Additionally, The Matlab Graphical User Interface Development Environment (GUIDE) simplifies the process of designing and building GUIs. GUIDE was also used to develop the interfaces to display the websites’ links before and after classification and the results obtained, from the RS algorithms and methods.

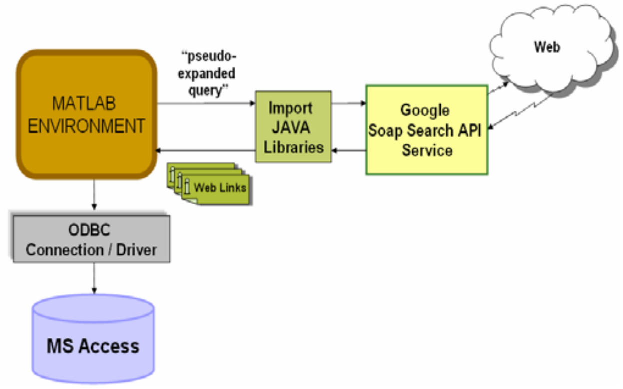


Fig. 5 Matlab Environment and Connections

A. Settings for Matlab-Web browser-Database connections

For this contribution, the selected *web browser* was GOOGLE. This browser is widely considered the strongest of the available online public browsers by the general population. Additionally, GOOGLE provides a set of routines, protocols and tools for building software applications such as the Application Program Interfaces (APIs) and Open Source Code for programmers. Some GOOGLE programming languages used include: Java, Perl, Python, PHP and .Net. The Google Soap Search API service allows software developers to query, directly, to WebPages. In order to access this service a license key is obtained in addition to creating a Google account, enabling to 1000 automatic queries per day. Matlab permits access to the Google Soap Search API service through Java library imports. Finally, the knowledge bases (KBs) are stored on MS Access. The database connection is based on the configuration of the Open Database Connectivity (ODBC) connection and ODBC-rmjdio driver. The ODBC User Data Source stores information about the connection to a specified data provider and the ODBC driver allows ODBC-enabled programs to get information from ODBC data sources. Fig. 5 illustrates the settings for this implementation, including the Matlab environment, the Internet connections and the Database connections employed.

B. Software Development

The search query can be created by allowing the user to indicate random terms in real time or by accessing the Category KB and extract a category with its subcategories. The process of classification of relevance is then applied to:

(1) *User query web search Interface*. This section allows users to instantly indicate a set of terms to search online. The web links obtained are classified using the Decision Process of Relevance based on Rough Sets (DPR-RS) be compared to the web links classified using Rough Sets. The resulting data is displayed to the user and it gives to the user the option to store it in the KB.

(2) *Category Web Search Interface*. Instead of receiving the query terms from the user, they are received from the Category KB. The system access to the KB, obtains the category and subcategories, and creates the search query. The search query is used to search online, obtain the corresponding weblinks, and classify those using DPR-RS. The classified links are stored into a knowledge base, named Weblinks KB.

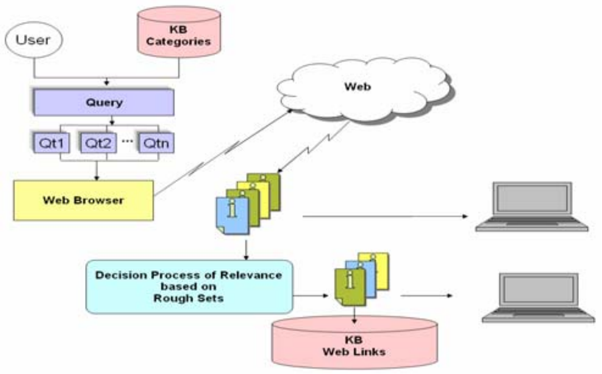


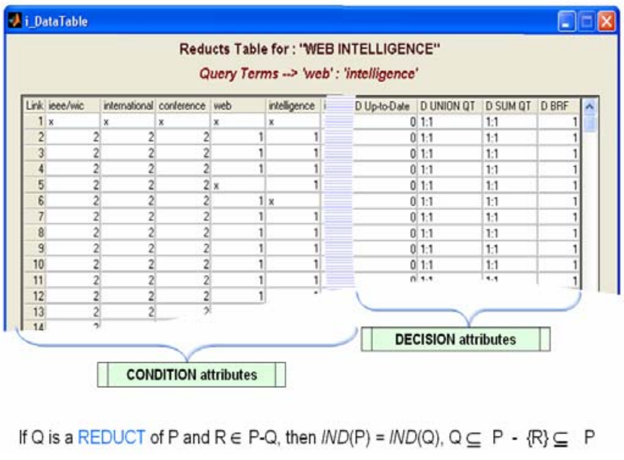
Fig. 6 Algorithm for the Classification model using the DPR-RS approach .

The Fig. 6 shows a graphical representation of the process from receiving the query terms until presenting the classified outputs (weblinks) in the display. The query terms are sent to the search engine, which returns a list of weblinks. Those weblinks are ranked by the web browser and the number of times that the user needs to go through the large list of websites to identify the sites of his/her interest. Also this list usually contains links that access to the same website. The weblinks is represented in the Fig. 6 by sheet with an 'i' from Internet. The set of web links goes to the DPR-RS engine to generate an ordered and shorter list of links, eliminating the repeated sites.

C. Decision Process of Relevance based on Rough Set Algorithm implementation

The Implemented RS algorithm, based upon Pawlak [14]. To explain the DPR-RS, a *search query* is selected as "Web Intelligence". These terms are sent to the *Search engine* and a set of weblinks are obtained. The terms included in the web *documents* are used to create the Information System or table, as previously mentioned. The first step is to create the Information System that is represented in an Initial Term Frequency Table (previously presented in Fig. 2). This table contains the total set of terms found in the web link *documents*. The decision attributes are the *UtD, Union QTF, Sum QTF,* and *BRF.*

The next step is to obtain the reducts of the e-dispensable table, which is based on the indiscernibility relation. Superfluous attributes are reduced and a new table is generated and can be viewed in Fig. 7.



Reducts Table for : "WEB INTELLIGENCE"

Query Terms --> 'web' : 'intelligence'

Link	ieee/wic	international	conference	web	intelligence	i	D Up-to-Date	D UNION QT	D SUM QT	D BRF
1	x	x	x	x	x		0	1:1	1:1	1
2	2	2	2	1	1		0	1:1	1:1	1
3	2	2	2	1	1		0	1:1	1:1	1
4	2	2	2	1	1		0	1:1	1:1	1
5	2	2	2	x	1		0	1:1	1:1	1
6	2	2	2	1	x		0	1:1	1:1	1
7	2	2	2	1	1		0	1:1	1:1	1
8	2	2	2	1	1		0	1:1	1:1	1
9	2	2	2	1	1		0	1:1	1:1	1
10	2	2	2	1	1		0	1:1	1:1	1
11	2	2	2	1	1		0	1:1	1:1	1
12	2	2	2	1						
13	2	2	2							
14	2									

CONDITION attributes

DECISION attributes

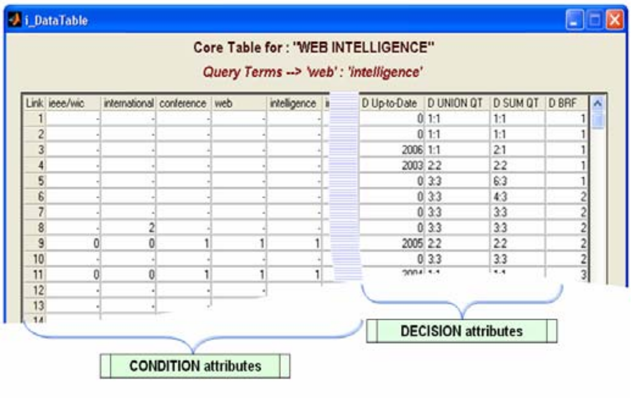
If Q is a REDUCT of P and R ∈ P-Q, then *IND*(P) = *IND*(Q), Q ⊆ P - {R} ⊆ P

Fig. 7 Table with the Reducts values

Having obtained the reducts for the table (information system), the core values are obtained and illustrated in Fig. 8.

The subsequent step obtains the rules for the information system. The websites are classified according those rules. The obtained list of weblinks are displayed to the user and stored in the Weblinks KB, as illustrated on Fig. 9.

These resultant websites are displayed in a web *Search display*, illustrated in Fig. 10. In the *Search display*, the user can create a new set of terms (web query terms) to search online. Each search obtains 10 links, therefore three "Times" would obtain the first 30 links indicates. In the same figure, the "Results" of this search are displayed in the panel, called "Results Web links ordered from Internet Browser (Google)". The first column enumerates the links, the second presents the *title* for the links, and the third contains their *URL*s. When the user "clicks" on any of the *titles* or *URL*s rows, a brief description (*document*) of the website is displayed in the *textbox* at the bottom of the panel.



i_DataTable

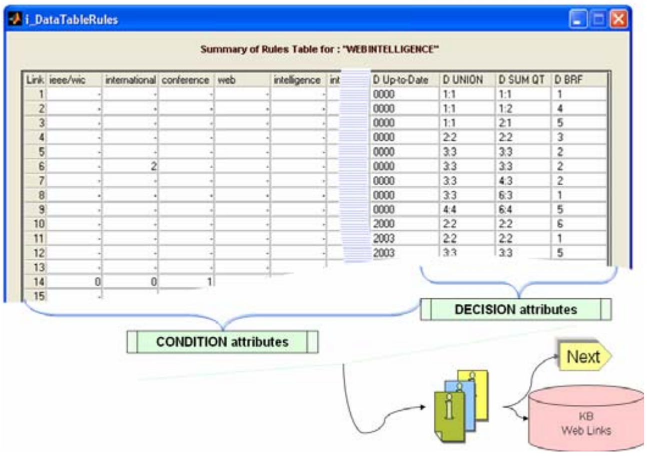
Core Table for : "WEB INTELLIGENCE"

Query Terms --> 'web' : 'intelligence'

Link	ieee/wic	international	conference	web	intelligence	i	D Up-to-Date	D UNION QT	D SUM QT	D BRF
1	-	-	-	-	-		0	1:1	1:1	1
2	-	-	-	-	-		0	1:1	1:1	1
3	-	-	-	-	-		2006	1:1	2:1	1
4	-	-	-	-	-		2003	2:2	2:2	1
5	-	-	-	-	-		0	3:3	6:3	1
6	-	-	-	-	-		0	3:3	4:3	2
7	-	-	-	-	-		0	3:3	3:3	2
8	-	2	-	-	-		0	3:3	3:3	2
9	0	0	1	1	1		2005	2:2	2:2	2
10	-	-	-	-	-		0	3:3	3:3	2
11	0	0	1	1	1		2004	1:1	1:1	3
12	-	-	-	-						
13	-	-								
14										

$$CORE(P) = \cap RED(P)$$

Fig. 8 Table with Core values



i_DataTableRules

Summary of Rules Table for : "WEBINTELLIGENCE"

Link	ieee/wic	international	conference	web	intelligence	int	D Up-to-Date	D UNION	D SUM QT	D BRF
1	-	-	-	-	-		0000	1:1	1:1	1
2	-	-	-	-	-		0000	1:1	1:2	4
3	-	-	-	-	-		0000	1:1	2:1	5
4	-	-	-	-	-		0000	2:2	2:2	3
5	-	-	-	-	-		0000	3:3	3:3	2
6	-	2	-	-	-		0000	3:3	3:3	2
7	-	-	-	-	-		0000	3:3	4:3	2
8	-	-	-	-	-		0000	3:3	6:3	1
9	-	-	-	-	-		0000	4:4	6:4	5
10	-	-	-	-	-		2000	2:2	2:2	6
11	-	-	-	-	-		2003	2:2	2:2	1
12	-	-	-	-	-		2003	3:3	3:3	5
13	-	-	-	-						
14	0	0	1							
15	-									

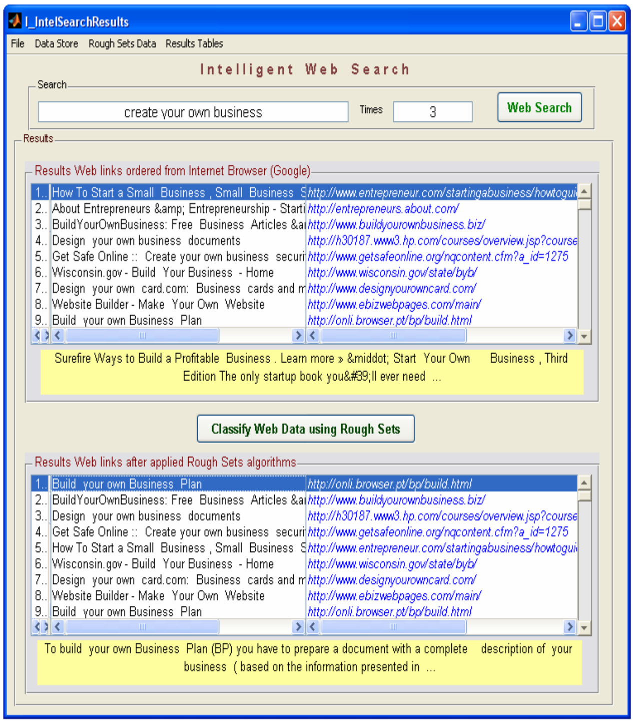
Fig. 9 Table with Core values

By pressing the button "Classify Web Data using Rough Set", those web links are classified according to their relevance to the categories based on rough sets. The portal asks the user to indicate the relevance method to classify, either by priority of decision attributes (PDA) or percentage of relevance (PRDA), as illustrated in Fig. 11.

Once the user selects the method of relevance for his/her preference, the classification procedure begins. The final classified weblinks are included in the panel "Results Web links after applied Rough Sets algorithms", included in Fig. 10.

V. Evaluation of System Performance

This section evaluates the performance of our system. As mentioned earlier, more attributes were involved in the definition of decision parameters such as *Up-to-Date*, frequency of the *Union of Query Terms* (*Union QTF*), frequency of the *Sum of Query Terms Frequency* (*Sum QTF*), and the *Blind Relevance Feedback* (*BRF*).



create your own business

Times 3

Results Web links ordered from Internet Browser (Google)

1..	How To Start a Small Business , Small Business S	http://www.entrepreneur.com/startingabusiness/howtogui
2..	About Entrepreneurs &amp; Entrepreneurship - Starti	http://entrepreneurs.about.com/
3..	BuildYourOwnBusiness: Free Business Articles &ai	http://www.buildyourownbusiness.biz/
4..	Design your own business documents	http://h30187.www3.hp.com/courses/overview.jsp?course
5..	Get Safe Online :: Create your own business securi	http://www.getsafeonline.org/nqcontent.cfm?a_id=1275
6..	Wisconsin.gov - Build Your Business - Home	http://www.wisconsin.gov/state/byb/
7..	Design your own card.com: Business cards and m	http://www.designyourowncard.com/
8..	Website Builder - Make Your Own Website	http://www.ebizwebpages.com/main/
9..	Build your own Business Plan	http://onli.browser.pt/bp/build.html

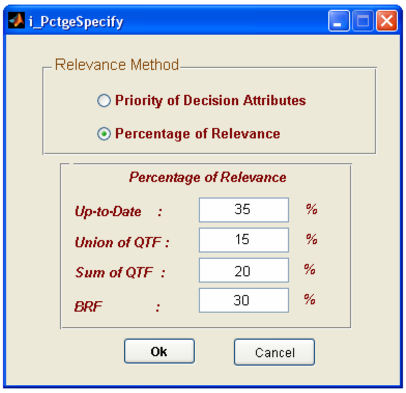
Surefire Ways to Build a Profitable Business . Learn more » &middot; Start Your Own Business , Third Edition The only startup book you&#39;ll ever need ...

Results Web links after applied Rough Sets algorithms

1..	Build your own Business Plan	http://onli.browser.pt/bp/build.html
2..	BuildYourOwnBusiness: Free Business Articles &ai	http://www.buildyourownbusiness.biz/
3..	Design your own business documents	http://h30187.www3.hp.com/courses/overview.jsp?course
4..	Get Safe Online :: Create your own business securi	http://www.getsafeonline.org/nqcontent.cfm?a_id=1275
5..	How To Start a Small Business , Small Business S	http://www.entrepreneur.com/startingabusiness/howtogui
6..	Wisconsin.gov - Build Your Business - Home	http://www.wisconsin.gov/state/byb/
7..	Design your own card.com: Business cards and m	http://www.designyourowncard.com/
8..	Website Builder - Make Your Own Website	http://www.ebizwebpages.com/main/
9..	Build your own Business Plan	http://onli.browser.pt/bp/build.html

To build your own Business Plan (BP) you have to prepare a document with a complete description of your business (based on the information presented in ...

Fig. 10 Final stage for the *Search display*



Relevance Method

○ Priority of Decision Attributes

● Percentage of Relevance

Percentage of Relevance

Up-to-Date :	35	%
Union of QTF :	15	%
Sum of QTF :	20	%
BRF :	30	%

Fig. 11 Choosing Reference Method in the *Priority display*

These attributes, together with the conditional attributes, are classified in the RS algorithm. In order to identify the relevance of websites, two methods based on the decision attributes were proposed in the previous section. These methods are PDA and PRDA. "Web Intelligence" *search query* is used to explain these methods. The number of links is 30 and the *search query* contains two terms: web and intelligence (Illustrated in Table II).

TABLE II

INDISCERNIBILITY PARTITION

Query:	Web Intelligence	
Total of links:	30	
Terms:	Term 1	Term 2
	web	intelligence

A. *Priority of Decision Attributes Method*

The PDA method consists of setting priorities to each of the decisional attributes. RS classifies the websites according to those attributes to distribute relevant links in priority sequence. The *Up-to-Date* decision attribute receives the higher priority; therefore the date of web publishing indicated in the website determines relevance. The second and third priorities consider the number of occurrences for all terms within the *search query* found in the "document". The last priority is given to BRF.

The *Union QTF* evaluates relevance for multiple term queries based upon a ratio of individual term occurrence, therefore in the case where a search with two terms (say *x* and *y*), a website with a term ratio of *x*:*y* = 3:3 will have a lower priority than one with a direct 7:7 ratio of occurrence.

The *Sum QTF* sets the relevance according to the sum of occurrences of the individual search terms in the web *document*. For example, a ratio of 2:9 (sum 11) will have higher priority than the ratio 3:5 (sum 8). When the sum of the ratios are the same, for example, 3:6 and 5:4 which both of their sum are 10, the priority is set higher to the first term and later to the second. The ratio 5:4 will have higher priority than 3:6 ratio of occurrence. These occurrences indicate how strongly the *document* is related to the query.

Finally, the fourth priority (*BRF*) indicates the relevance based on the website rank provided by the *web browser*. Consequently, the first links have higher rank of *BRF* while the last links have lower rank. The Priority of Decision Attributes method is exemplified in Fig. 12.

After analyzing the results, it can be concluded that the *Up-to-Date* indices seem to be present in certain groups of web links, such as conferences, journals, and books; reducing the time to identify relevant information of this group. The results illustrated in Fig. 13 demonstrate that the first two links have the most recent *Up-to-Date* time indices, with a *Union QTF* and *Sum QTF* of ratio occurrences of one and three, respectively, for the first links and one and two (respectively) for the second link. Links number 3 and 19 demonstrate that many links must be scrutinized by the user before the two most relevant links are found by the web browser. This method saves time as the relevant links are assigned the highest priority for immediate access, reducing, significantly, the required time to identify them.

Studying records number 12 to 19, however, the websites' "documents" contain data with high levels of priority on decision attributes other than the *Up-to-Date* attribute. To address this discrepancy in relevance, another method to approach the relevance of websites is proposed. This method assigns a degree of relevance (percentage value) to each decision attributes.

B. *Percentage of Relevance of Decision Attributes Method*

While some groups of web links seem to be strongly favored to the Priority of Decision Attributes method, some other groups require a different approach to set the relevance base upon the terms contained.

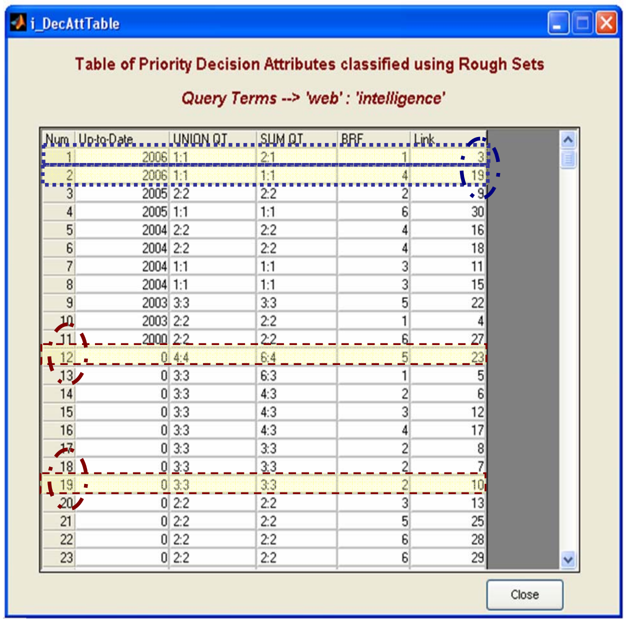
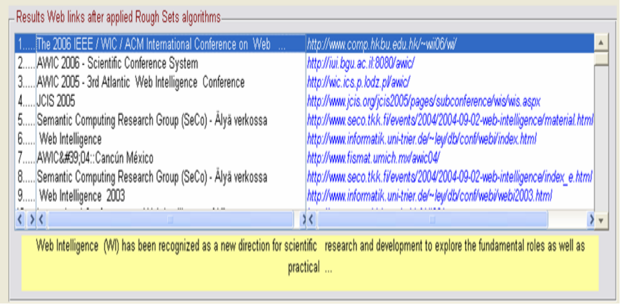


Table of Priority Decision Attributes classified using Rough Sets

Query Terms --> 'web' : 'intelligence'

Num	Up-to-Date	UNION QT	SUM QT	BRF	Link
1	2006	1:1	2:1	1	3
2	2006	1:1	1:1	4	19
3	2005	2:2	2:2	2	9
4	2005	1:1	1:1	6	30
5	2004	2:2	2:2	4	16
6	2004	2:2	2:2	4	18
7	2004	1:1	1:1	3	11
8	2004	1:1	1:1	3	15
9	2003	3:3	3:3	5	22
10	2003	2:2	2:2	1	4
11	2000	2:2	2:2	6	27
12	0	4:4	6:4	5	23
13	0	3:3	6:3	1	5
14	0	3:3	4:3	2	6
15	0	3:3	4:3	3	12
16	0	3:3	4:3	4	17
17	0	3:3	3:3	2	8
18	0	3:3	3:3	2	7
19	0	3:3	3:3	2	10
20	0	2:2	2:2	3	13
21	0	2:2	2:2	5	25
22	0	2:2	2:2	6	28
23	0	2:2	2:2	6	29

Fig. 12 Table for Priority of Decision Attributes using RS Results



Results Web links after applied Rough Sets algorithms

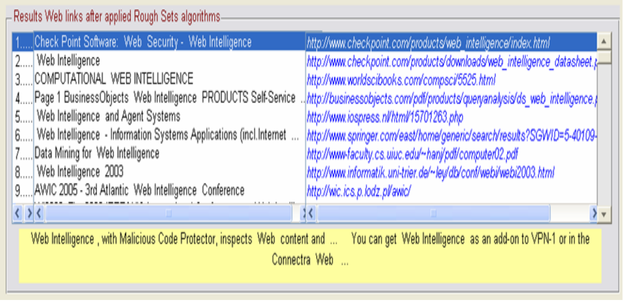
1.....The 2006 IEEE / WIC / ACM International Conference on Web ...	http://www.comp.hkbu.edu.hk/~wii06/wi/
2.....AWIC 2006 - Scientific Conference System	http://iui.bgu.ac.il:8080/awic/
3.....AWIC 2005 - 3rd Atlantic Web Intelligence Conference	http://wic.ics.p.lodz.pl/awic/
4.....JCIS 2005	http://www.jcis.org/jcis2005/pages/subconference/wis/wis.aspx
5.....Semantic Computing Research Group (SeCo) - Älyä verkossa	http://www.seco.tkk.fi/events/2004/2004-09-02-web-intelligence/material.html
6..... Web Intelligence	http://www.informatik.uni-trier.de/~ley/db/conf/webi/index.html
7.....AWIC&#39;04::Cancún México	http://www.fismat.umich.mx/awic04/
8.....Semantic Computing Research Group (SeCo) - Älyä verkossa	http://www.seco.tkk.fi/events/2004/2004-09-02-web-intelligence/index_e.html
9..... Web Intelligence 2003	http://www.informatik.uni-trier.de/~ley/db/conf/webi/webi2003.html

Web Intelligence (WI) has been recognized as a new direction for scientific research and development to explore the fundamental roles as well as practical ...

Fig. 13 Links classified by PDA method

This method consists of assigning a level of priority (percentage) to each of the decision attributes. The percentage assigned will indicate the level of relevance for an attribute of interest during the RS classification process. This priority assignment can dynamically indicate which decision attribute with higher priority is.

The assignment of percentage of relevance can be dynamically updated. Our first example, sets the higher percentage to the *Union QTF*, giving with this more relevance to the terms contained. The percentage assigned to the *Up-to-Date* and *BRF* is 10% of the total percentage of relevance; whereas the decision attributes for *Union QTF* is 60% and *Sum QTF* is 20%. Therefore, the weblinks listed will be highly related to the search terms in the query, providing an immediate access to groups of links that fulfill the requirements. For examples, certain classes of the categories, such as "products" and "merchandises", have great value for some categories.



Results Web links after applied Rough Sets algorithms

1.....	Check Point Software: Web Security - Web Intelligence	http://www.checkpoint.com/products/web_intelligence/index.html
2.....	Web Intelligence	http://www.checkpoint.com/products/downloads/web_intelligence_datasheet.p
3.....	COMPUTATIONAL WEB INTELLIGENCE	http://www.worldscibooks.com/compsci/5525.html
4.....	Page 1 BusinessObjects Web Intelligence PRODUCTS Self-Service ...	http://businessobjects.com/pdf/products/queryanalysis/ds_web_intelligence.p
5.....	Web Intelligence and Agent Systems	http://www.iospress.nl/html/15701263.php
6.....	Web Intelligence - Information Systems Applications (incl.Internet ...	http://www.springer.com/east/home/generic/search/results?SGWID=5-40109-
7.....	Data Mining for Web Intelligence	http://www-faculty.cs.uiuc.edu/~hanj/pdf/computer02.pdf
8.....	Web Intelligence 2003	http://www.informatik.uni-trier.de/~ley/db/conf/webi/webi2003.html
9.....	AWIC 2005 - 3rd Atlantic Web Intelligence Conference	http://wic.ics.p.lodz.pl/awic/

Web Intelligence , with Malicious Code Protector, inspects Web content and ... You can get Web Intelligence as an add-on to VPN-1 or in the Connectra Web ...

Fig. 14 Links classified by PRDA method using RS

Continuing with the initial *search query* ("Web Intelligence"), the obtained results for this method are illustrated in Fig. 14 and the table with the Percentage of Relevance can be viewed on Fig. 15.

TABLE IV
FIRST RANKED LINK BY THE WEB BROWSER

Title:	Web Intelligence Consortium
URL:	http://wi-consortium.org
Document:	Empty snippet

As mentioned earlier, "document" or snippet is provided by the *search engine*. Therefore, in the case where *document* is empty, the proper classification of this site is partially reduced. The results obtained in our priority method indicates that this link only contains values in the *Union QTF* and *Sum QTF* of the terms occurrences, and BRF of the Decision Attributes of 1, 2, and 4, respectively. In comparison, the first obtained link after classifying using RS and based on our Priority (PDA) method is illustrated in Table V.

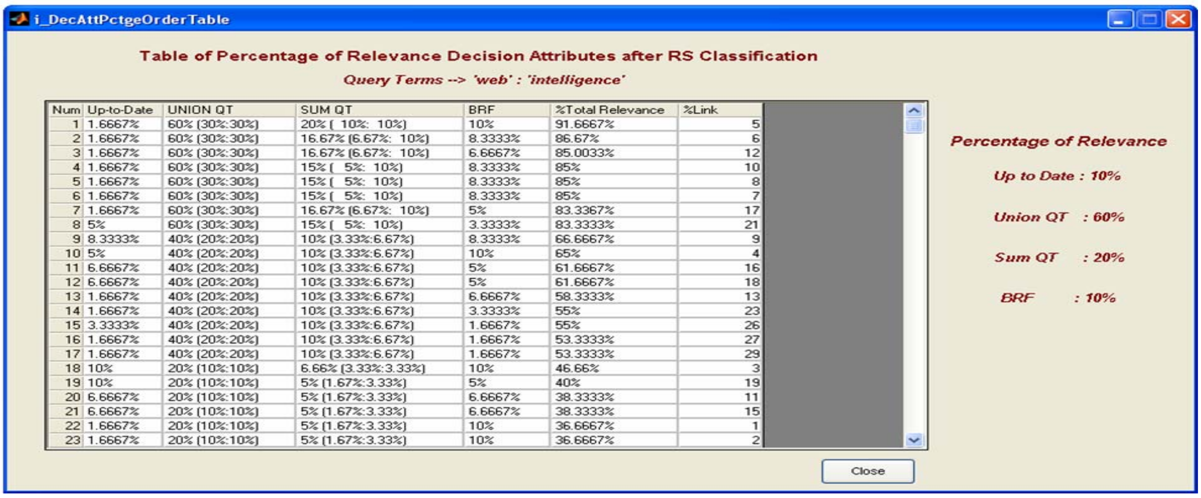


Table of Percentage of Relevance Decision Attributes after RS Classification

Query Terms --> 'web' : 'intelligence'

Num	Up-to-Date	UNION QT	SUM QT	BRF	%Total Relevance	%Link
1	1.6667%	60% (30%:30%)	20% (10%: 10%)	10%	91.6667%	5
2	1.6667%	60% (30%:30%)	16.67% (6.67%: 10%)	8.3333%	86.67%	6
3	1.6667%	60% (30%:30%)	16.67% (6.67%: 10%)	6.6667%	85.0033%	12
4	1.6667%	60% (30%:30%)	15% (5%: 10%)	8.3333%	85%	10
5	1.6667%	60% (30%:30%)	15% (5%: 10%)	8.3333%	85%	8
6	1.6667%	60% (30%:30%)	15% (5%: 10%)	8.3333%	85%	7
7	1.6667%	60% (30%:30%)	16.67% (6.67%: 10%)	5%	83.3367%	17
8	5%	60% (30%:30%)	15% (5%: 10%)	3.3333%	83.3333%	21
9	8.3333%	40% (20%:20%)	10% (3.33%:6.67%)	8.3333%	66.6667%	9
10	5%	40% (20%:20%)	10% (3.33%:6.67%)	10%	65%	4
11	6.6667%	40% (20%:20%)	10% (3.33%:6.67%)	5%	61.6667%	16
12	6.6667%	40% (20%:20%)	10% (3.33%:6.67%)	5%	61.6667%	18
13	1.6667%	40% (20%:20%)	10% (3.33%:6.67%)	6.6667%	58.3333%	13
14	1.6667%	40% (20%:20%)	10% (3.33%:6.67%)	3.3333%	55%	23
15	3.3333%	40% (20%:20%)	10% (3.33%:6.67%)	1.6667%	55%	26
16	1.6667%	40% (20%:20%)	10% (3.33%:6.67%)	1.6667%	53.3333%	27
17	1.6667%	40% (20%:20%)	10% (3.33%:6.67%)	1.6667%	53.3333%	29
18	10%	20% (10%:10%)	6.66% (3.33%:3.33%)	10%	46.66%	3
19	10%	20% (10%:10%)	5% (1.67%:3.33%)	5%	40%	19
20	6.6667%	20% (10%:10%)	5% (1.67%:3.33%)	6.6667%	38.3333%	11
21	6.6667%	20% (10%:10%)	5% (1.67%:3.33%)	6.6667%	38.3333%	15
22	1.6667%	20% (10%:10%)	5% (1.67%:3.33%)	10%	36.6667%	1
23	1.6667%	20% (10%:10%)	5% (1.67%:3.33%)	10%	36.6667%	2

Percentage of Relevance

Up to Date : 10%

Union QT : 60%

Sum QT : 20%

BRF : 10%

Fig. 15 Table Percentage of Relevance Decision Attributes using RS

VI. ANALYSIS OF RESULTS

For the example "web intelligence" *search query*, the total of links processed is 30 and the terms are (illustrated in Table III):

TABLE III
SEARCH QUERY EXAMPLE

Query:	Web Intelligence	
Total of links:	30	
Terms:	Term 1	Term 2
	web	intelligence

Those terms are sent to the *search engine* to obtain the link from the web. The obtained web links indicate that the first obtained link ranked by the *web browser* is (illustrated in Table IV:

TABLE V
FIRST RANKED LINK AFTER CLASSIFYING USING PDA METHOD

Title:	The 2006 IEEE/ WIC / ACM International Conference on Web …
URL:	http://www.comp.hkbu.edu.hk/~wii06/wi/
Document:	Web Intelligence (WI) has been recognized as a new direction for scientific research and development to explore the fundamental roles as well as practical …

According to our priority method, Table VI presents the information obtained for this link.

TABLE VI
INFORMATION FOR THE FIRST RANKED LINK USING PDA METHOD

Up-to-Date	*Union QTF*	*Sum QTF*	*BRF*
2006	1:1	2:1	3

Finally, comparing the first obtained link after classification using RS based on our percentage of relevancies (PRDA) method, the links ranked is according to the terms contained.

The Table VI shows this link.

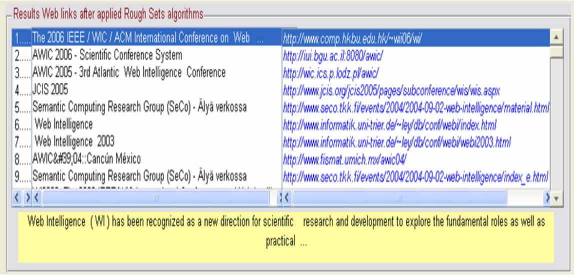
TABLE VII
FIRST RANKED LINK AFTER CLASSIFYING USING PDA METHOD

Title:	Check Point Software: Web Security – Web Intelligence
URL:	http://www.checkpoint.com/products/web_intelligence/index.html
Document:	Web Intelligence, with Malicious Code Protector, inspects Web content and … You can get Web Intelligence as an add-on to VPN-1 or in the Connectra Web

Fig 15 contains the Percentage of Relevance of Decision attributes, which indicates the following values, included in Table VIII:

TABLE VIII
INFORMATION FOR THE FIRST RANKED LINK USING PRDA METHOD

% *Up-to-Date*	% *Union QTF*	% *Sum QTF*	% *BRF*	% *Relevance*	*Link*
1.6667%	60% (30%:30%)	20% (10%:10%)	10%	91.667%	5



Results Web links after applied Rough Sets algorithms

1.....The 2006 IEEE / WIC / ACM International Conference on Web ...	http://www.comp.hkbu.edu.hk/~wii06/wi/
2.....AWIC 2006 - Scientific Conference System	http://iui.bgu.ac.il:8080/awic/
3.....AWIC 2005 - 3rd Atlantic Web Intelligence Conference	http://wic.ics.p.lodz.pl/awic/
4.....JCIS 2005	http://www.jcis.org/jcis2005/pages/subconference/wis/wis.aspx
5.....Semantic Computing Research Group (SeCo) - Älyä verkossa	http://www.seco.tkk.fi/events/2004/2004-09-02-web-intelligence/material.html
6..... Web Intelligence	http://www.informatik.uni-trier.de/~ley/db/conf/webi/index.html
7..... Web Intelligence 2003	http://www.informatik.uni-trier.de/~ley/db/conf/webi/webi2003.html
8.....AWIC&#39;04::Cancún México	http://www.fismat.umich.mx/awic04/
9.....Semantic Computing Research Group (SeCo) - Älyä verkossa	http://www.seco.tkk.fi/events/2004/2004-09-02-web-intelligence/index_e.html

Web Intelligence (WI) has been recognized as a new direction for scientific research and development to explore the fundamental roles as well as practical ...

Fig. 16 Links classified by PRDA method using RS

Up-to-Date and *BRF* have a 10% of relevance, while both the *Union QTF* and *Sum QTF* have a 60% and 20% of relevance. This example uses the same *search query*; however, it allows to help the decision attribute of interests. The following example varies the percentage of relevance and sets *Up-to-Date* to 70%, *Union QTF* to 20%, *Sum QTF* to 5%,and *BRF* to 5%. The classified weblinks obtained are located in priority sequence, which have been created more recently. Fig 16 contains the weblinks classified under these conditions and Fig 17 illustrates the percentages of relevance for each link.

It can be concluded that indicating a desired relevance (through percentage), will identify relevant web pages according to a set of terms. Thus, the percentage values will be adjusted in order to indicate the priority for each decision attribute. For this reason, this second proposed method was chosen to classify the web links through rough sets.

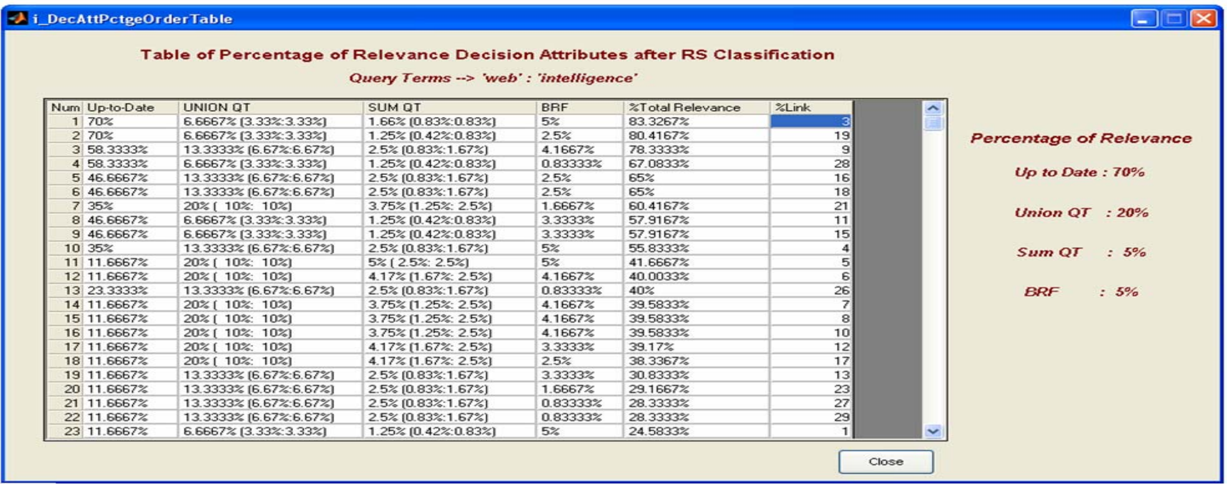


Table of Percentage of Relevance Decision Attributes after RS Classification
Query Terms --> 'web' : 'intelligence'

Num	Up-to-Date	UNION QT	SUM QT	BRF	%Total Relevance	%Link
1	70%	6.6667% (3.33%:3.33%)	1.66% (0.83%:0.83%)	5%	83.3267%	3
2	70%	6.6667% (3.33%:3.33%)	1.25% (0.42%:0.83%)	2.5%	80.4167%	19
3	58.3333%	13.3333% (6.67%:6.67%)	2.5% (0.83%:1.67%)	4.1667%	78.3333%	9
4	58.3333%	6.6667% (3.33%:3.33%)	1.25% (0.42%:0.83%)	0.83333%	67.0833%	28
5	46.6667%	13.3333% (6.67%:6.67%)	2.5% (0.83%:1.67%)	2.5%	65%	16
6	46.6667%	13.3333% (6.67%:6.67%)	2.5% (0.83%:1.67%)	2.5%	65%	18
7	35%	20% (10%: 10%)	3.75% (1.25%: 2.5%)	1.6667%	60.4167%	21
8	46.6667%	6.6667% (3.33%:3.33%)	1.25% (0.42%:0.83%)	3.3333%	57.9167%	11
9	46.6667%	6.6667% (3.33%:3.33%)	1.25% (0.42%:0.83%)	3.3333%	57.9167%	15
10	35%	13.3333% (6.67%:6.67%)	2.5% (0.83%:1.67%)	5%	55.8333%	4
11	11.6667%	20% (10%: 10%)	5% (2.5%: 2.5%)	5%	41.6667%	5
12	11.6667%	20% (10%: 10%)	4.17% (1.67%: 2.5%)	4.1667%	40.0033%	6
13	23.3333%	13.3333% (6.67%:6.67%)	2.5% (0.83%:1.67%)	0.83333%	40%	26
14	11.6667%	20% (10%: 10%)	3.75% (1.25%: 2.5%)	4.1667%	39.5833%	7
15	11.6667%	20% (10%: 10%)	3.75% (1.25%: 2.5%)	4.1667%	39.5833%	8
16	11.6667%	20% (10%: 10%)	3.75% (1.25%: 2.5%)	4.1667%	39.5833%	10
17	11.6667%	20% (10%: 10%)	4.17% (1.67%: 2.5%)	3.3333%	39.17%	12
18	11.6667%	20% (10%: 10%)	4.17% (1.67%: 2.5%)	2.5%	38.3367%	17
19	11.6667%	13.3333% (6.67%:6.67%)	2.5% (0.83%:1.67%)	3.3333%	30.8333%	13
20	11.6667%	13.3333% (6.67%:6.67%)	2.5% (0.83%:1.67%)	1.6667%	29.1667%	23
21	11.6667%	13.3333% (6.67%:6.67%)	2.5% (0.83%:1.67%)	0.83333%	28.3333%	27
22	11.6667%	13.3333% (6.67%:6.67%)	2.5% (0.83%:1.67%)	0.83333%	28.3333%	29
23	11.6667%	6.6667% (3.33%:3.33%)	1.25% (0.42%:0.83%)	5%	24.5833%	1

Percentage of Relevance
Up to Date : 70%
Union QT : 20%
Sum QT : 5%
BRF : 5%

Fig. 17 Table Percentage of Relevance Decision Attributes using RS

VII. CONCLUSIONS

A new approach to reduce web search times using rough sets is proposed. The decision attributes defined to identify priority of web links are: *Up-to-Date*, *Union QTF*, *Sum QTF*, and *BRF* attributes. The approach involved two methods for defining the priority of web links. These methods were the Priority of Decision Attributes (PDA) and the Percentage of Relevance of Decision Attributes (PRDA). After evaluating these methods, PRDA demonstrated better results in its ability to vary relevance of web links by adjusting the percentage of priority or relevance for each decision attribute. By adjusting these percentages, the PRDA method sets higher priority in order to support more recent web links or to favor the ratio (occurrences) of query terms in the “document”. Moreover, it can also indicate the level of influence of the *BRF* to the relevance classification of web data. After applying rough sets, the obtained web data assures a list of web links classified in priority sequence. The primary relevant web links are located for instant access, reducing the time for web searches. Consequently, the implementation of this approach automatically generates and updates the *knowledge base,* which contains the relevant web links related to the user *search query* in priority sequence.

ACKNOWLEDGMENT

The authors would like to thank Luis H. Garcia Islas for proofreading the revised version of this Paper.

REFERENCES

[1] Boose, J.H. and Gaines, B.R., Knowledge acquisition tools for expert systems**.** London; Toronto : Academic Press, c1988. ISBN. 0122732510 (vol.2).

[2] Chang, C.-H. and Hsu, C.-C., Enabling Concept-Based Relevance Feedback for Information Retrieval on the WWW *IEEE Transactions on Knowledge and Data Engineering*, vol. 11, pp. 595-609, 1999.

[3] Chekuri, C., Goldwasser, M. H., Raghavan, P., and Upfal, E., Web search using automatic classification *In Proc. of the 6th International World Wide Web Conference (WWW)*, vol. 1997.

[4] De Cock, M. and Cornelis, C., Fuzzy Rough Set Based Web Query Expansion *in: Proceedings of Rough Sets and Soft Computing in Intelligent Agent and Web Technology, International Workshop at WIIAT2005 (2005 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology)*, p. 9-16, 2005

[5] Frigui, H. and Nasraoui, F., A fast algorithm for discovering categories and attribute relevance in web data *In Proc on Fuzzy Information Processing Society (NAFIPS. 2002) Annual Meeting of the North American*, vol. pp. 280-285, Jun, 2002.

[6] He, D., A Study of Self-Organizing Map in Interactive Relevance Feedback 3rd *Intl. Conf. on Information Technology: New Generations (ITNG 2006)*, vol. pp. 394-401, Apr, 2006.

[7] Hiemstra, D. and Robertson, S., Relevance Feedback for Best Match Term Weighting Algorithms in Information Retrieval *In A. F. Smeaton and J. Callan, editors, Proceedings of the Joint DELOS-NSF Workshop on Personalisation and Recommender Systems in Digital Libraries*, vol. pp. 37-42, Jun, 2001.

[8] Jang J.-S. R., Sun C.-T. Mizutani E. Neuro-Fuzzy and Soft Computing: A computational approach to learning and machine intelligence. Matlab Curriculum Series. Edit. Prentice Hall. 1997.

[9] Komorowski, J., Pawlak, Z., Polkowski, L., and Skowron, A., Rough Sets: A Tutorial *In: Pal, S.K., Skowron, A. (eds.): Rough Fuzzy Hybridization - A New Trend in Decision-Making* , vol. pp. 3-98, 1999.

[10] Lingras P. Rough set clustering for web mining *In Proceedings of 2002 World Congress on Computational Intelligence, IEEE International Conference on Fuzzy Systems (FUZZ-IEEE(02) Special Session on Computational Web Intelligence (CWI)*, vol. pp. 1039-44, 2002.

[11] Martinez E. Mayorga R. V., "An Architecture for the Coupling of Intelligent Computer Interfaces with Intelligent Systems: An Online Internet Portals Customization Application. Proceedings 4th *ANIROB/IEEE-RAS* Intl. Symposium on Robotics and Automation, Queretaro, Mexico, August, 25-27, 2004

[12] Mock, K., Dynamic email organization via relevance categories *In Proc. 11th Intl. Conf. on Tools with Artificial Intelligence*, vol. pp. 399-405, Nov, 1999.

[13] Ngo C. L. and Nguyen, H. S., A method of web search result clustering based on rough sets *Proceedings of 2005 IEEE/WIC/ACM International Conference on Web Intelligence*, vol. pp. 673-679, Sep, 2005.

[14] Pawlak, Z.. Rough sets : theoretical aspects of reasoning about data. Dordrecht ; Boston : Kluwer Academic Publishers, c1991. ISBN. 0792314727 (HB : acid free paper)

[15] Rojanavasu, P., Pinngern, 0. Extended Rough Fuzzy Sets For Web Search Agent, Proceedings of the 25'h International Conference on Information Technology Interfaces June 16-19, 2003, Cavtat, Croatia pp. 403-407

[16] Rui, Y. and Huang, T. S., A novel relevance feedback technique in image retrieval *In Proc. ACM Multimedia* , vol. pp. 67-70, 1999.

[17] Shen, D., Chen, Z., Yang, Q., Zeng, H.-J., Zhang, B., Lu, Y., and Ma, W.-Y., Web-page classification through summarization *In Proc. of the 27th annual international conference on Research and development in information retrieval*, vol. pp. 242-249, 2004.

[18] Takama, Y., Consideration of Relevance Feedback on Keyword Space for Interactive Information Retrieval *IEEE Conference on Cybernetics and Intelligent Systems (CIS2004)*, vol. 1, pp. 324-328, 2004.

[19] Wu, Y. and Zhang, A., An Adaptive Classification Method for Multimedia Retrieval *"", in IEEE International Conference on Multimedia and Expo (ICME'03)*, vol. pp. 757-760, Jul, 2003.

[20] Yi, G., Hu, H., and Lu, H., Web Document Classification Based on Extended Rough set *Sixth International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT 2005)*, vol. pp. 916-919, Dec, 2005.

Erika Martinez Ramirez, obtained her B.Sc. in Administrative Computer Systems from University of Celaya, Mexico; and her M.Sc. in Instrumentation and Automatic Control from the Autonomous University of Queretaro, Mexico. She is currently working in the Information Technology Department, at Agriculture Financial Service Corporation, Lacombe, Alberta, Canada. She has worked in projects involving Fuzzy Inference Systems, Artificial Neural Networks, and Rough Sets Theory. Her research interests include Soft Computing, Human-Computer Interaction, Human-Computer-Automated Systems, Personalization /Customization, Web Intelligence and application of AI to Engineering problems.

Dr. Rene V. Mayorga, is the head of the *W.I.S.E* (Wise & Intelligent Systems & Entities) Lab at the University of Regina. He is interested in the development of *Artificial / Computational Sapience [Wisdom]* (as an extension of Artificial / Computational Intelligence and Soft Computing) and *MetaBotics* (as a generalization of Robotics) as new disciplines. He is involved in the development of Paradigms for Intelligent and *Sapient [Wise]* Systems, *Sapient* Decision & Control, and *MetaBots*; and their application to Robotics (Robot and Multi-Robot Systems Motion Planning and Design), Automation, Manufacturing (Decision Making, Optimization), Human-Computer Interaction/Interface, and Software Engineering. Dr. Mayorga has recently co-edited the Book: *Toward Artificial Sapience, Principles and Methods for Wise Systems,* published by Springer in December 2007. Dr. Mayorga is the Editor in Chief of the journal of *Applied Bionics and Biomechanics*; and Associate Editor for: the *Journal of Control and Intelligent Systems,* the *International Journal of Robotics and Automation*, the *Journal Intelligent Service Robotics,* and the *Journal of Robotics.* He has been in several occasions the General Chair of the *ANIROB International Symposium of Robotics and Automation* and Editor of the corresponding Proceedings.