Use of Semantic Relatedness for Intelligent Access to Cultural Information

Harkiran Kaur
Thapar University,Patiala
Deptt. Of Comp. Sc. & Engg.
Thapar University, Patiala

Seema Bawa, PhD.
Thapar University,Patiala
Deptt. Of Comp. Sc. & Engg.
Thapar University, Patiala

ABSTRACT
This paper proposes the idea of providing intelligent access to cultural information available on a cultural portal (developed for the purpose of enhancing online access to Sikh cultural heritage. This paper proposes two kinds of search – Page Rank wise search and searching metadata. This search facility makes the system intelligent enough to understand the information required by the user. Links of pages relevant to the search are made available to the user (Page Rank wise); the most relevant result at first position, next relevant result at the second position and so on. Searching metadata helps the user to know about the information available on website and contribute new cultural information.

Keywords
semantics; semantic relatedness; intelligent access; cultural information; semantic search.

1. INTRODUCTION
One of the important focus areas of Cultural Computing is providing intelligent access to cultural information. [3]

This paper focuses on the need of Semantic search facility on the cultural portal. Semantic Relatedness focuses on the likeness of meaning of the words in consideration. If the two words have same meaning, then on searching about any of the two should produce same results. Otherwise, if a user searches one word or the other, the system will produce different results, hindering the search mechanism. [2]

Using Semantic relatedness for word search has various applications. The Semantics are well understood by humans, than computers. [4] To make the computers understand the semantics, some of the approaches have been proposed in this paper. It is kind of artificial intelligence system, to make the system intelligent enough to understand the meanings of words. Semantic relatedness of words is helpful in determining the page rank (for the desired information). A set of words can be represented or written in many ways. [7]

2. CASE STUDY
2.1 Implementing Semantic Relatedness in search facility on Cultural Portal
Suppose, a user searches for (say) five takhats of Sikhs, on the cultural portal. So, this can be searched by using any of the following set of words:

i. five takhats
ii. 5 takhts
iii. 5 takhats
iv. panj takhat (Panj means 5)
v. Five takhats of Sikhs,
vi. punj takhat
vii. takhats of Sikhs,

and many more.

The task of this approach is to make the system intelligent enough to understand that - any of these, means the same thing.

Above example shows, the difference among all the words above (1-6) is very minor. These words differ from each other, just in case of spellings and orientation of words. Also, '5' is represented as 'panj/punj' in Punjabi and 'takhat' can be written both ways, that is, 'takht/takhats'.

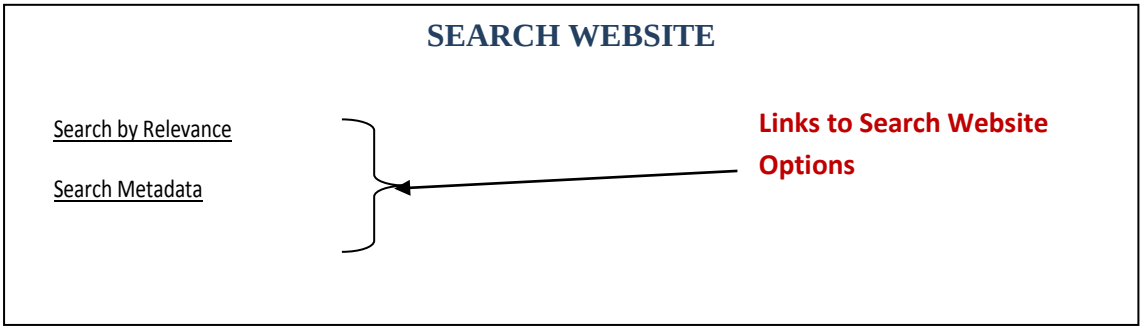


Fig 1: Search Portal Options

Despite small changes in spellings or orientation of words, words can also be represented in Romans, as in earlier example.

Example – 'panj pyare' is a word in Punjabi which can be written in roman like this. But, in English, it is interpreted as 'five beloved ones'. That is, words can be written differently using different languages.

The need of the hour is to understand the meaning of the word searched by the user, despite of the differences in case of spellings, orientation of words or the entirely different way of representing a particular word.

To make the search website feature complete in itself, this feature is provided in two kinds – Search by relevance (English and Punjabi) and Searching Metadata.

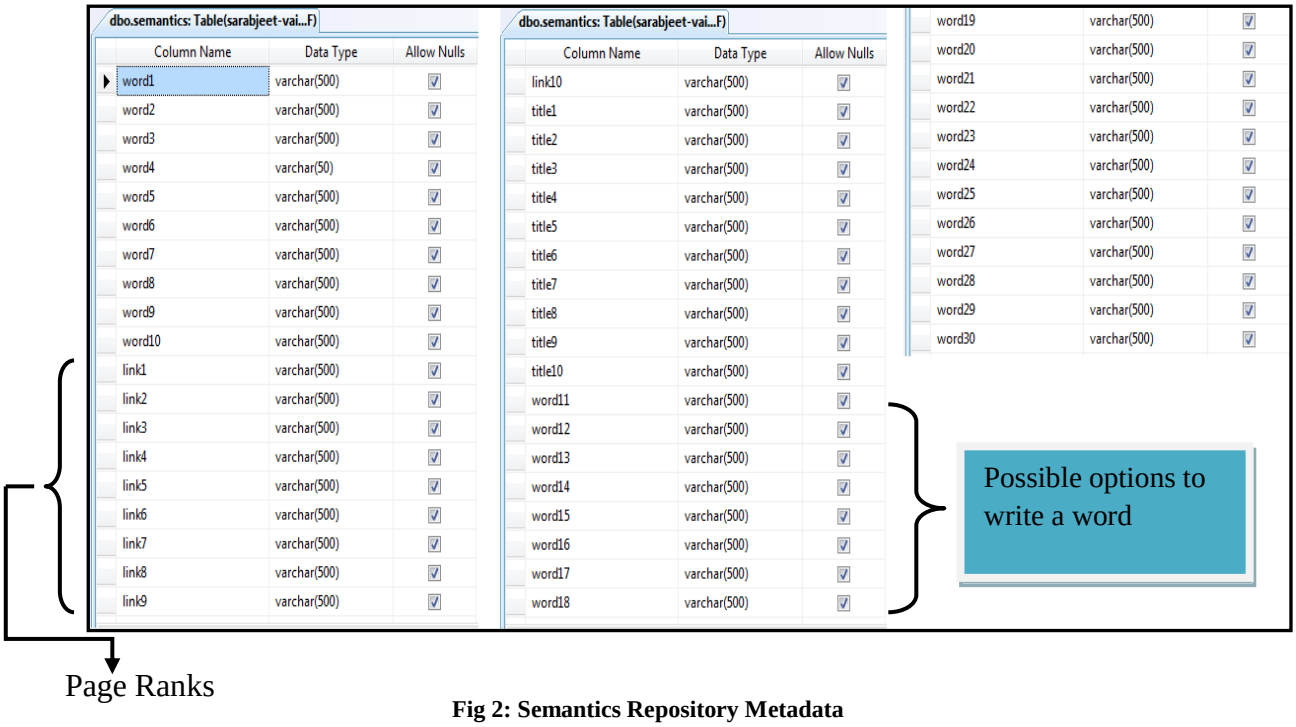


Fig 2: Semantics Repository Metadata

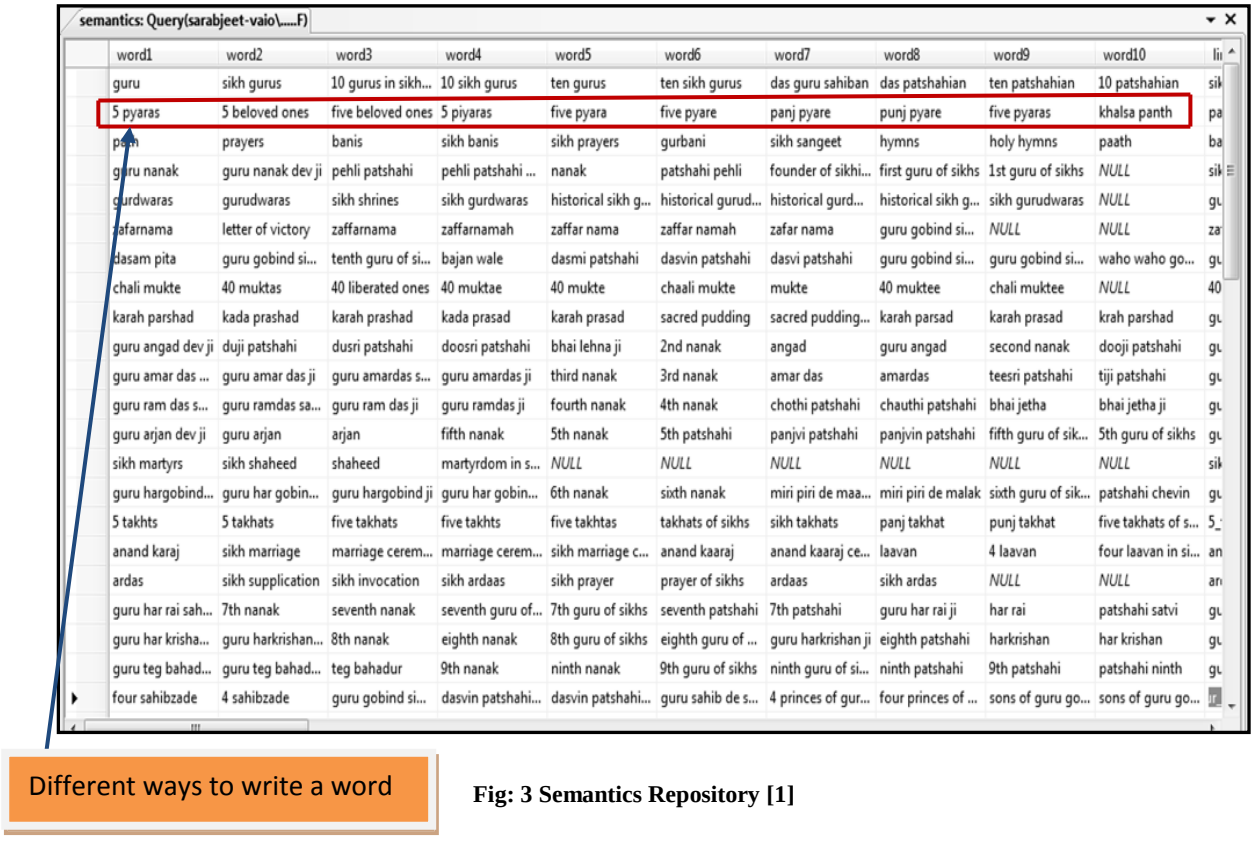


Fig: 3 Semantics Repository [1]

3. COMPARISON OF PAGE RANKING ALGORITHMS

Table 1: Comparison of various Page Ranking Algorithms [8] – [9]

	Parameters	**Algorithms**			
		Page Rank	**Distance Rank**	**HITS Algorithm**	**Dirichlet Rank**
1.	Working Method	Scores are computed at index time and results are arranged with respect to important pages for user	Minimum average distance between pages is calculated to compute the scores	Computes scores of pages that are much relevant to the query of user	Same as page rank but the transition probabilities are computed using Bayesian estimation
2.	Web Mining Technique	Web Structure Mining (WSM)	Web Structure Mining (WSM)	Both Web Structure Mining (WSM) and Web Content Mining (WCM)	Web Structure Mining (WSM)
3.	Implementation in which Search Engine?	Google search engine	Research Model	IBM prototype	Research Model
4.	Relevance	Less relevant as it ranks the web pages at indexed time	More relevant as it calculates the minimum average distance	More relevant as it considers content in its input parameters	More relevant than page rank algorithm as it calculates the transition probabilities using Bayesian estimation
5.	Input Parameters for algorithm	Back links	Back links	Content, Forward links and back links	Back links
6.	Complexity of algorithm	O (log N)	O (log N)	< O (log N)	O (log N)
7.	Quality of Results	Medium	Better quality of results than page rank algorithm	Low quality of results than page rank algorithm	More effective results
8.	Algorithm Importance	Much Important as back links are considered as input parameters	Much important as back links act as the input parameters.	Moderate importance as Hub and authorities scores is used	Stability is the important factor to make it more reliable
9.	Limitation	This algorithm is query independent.	This algorithm needs to work along with the page rank.	There is a problem of efficiency and topic drift.	Dirichlet rank algorithm needs to work along with page rank.

4. PAGE RANK in SEARCHING

Also. The system should provide facility to generate more relevant results, that is, page rank wise, depending upon the set of words searched for. Page rank wise results means, the most relevant result – at the first position, then the next near relevant result and so on. Page rank determines the relative importance of a search result in a set of search results.

4.1 Search by Relevance (English) on the Cultural Portal

It allows the user to type a set of words about the information required by them.
Different options of a particular word are stored in the form of repository shown in Figure 2 and Figure 3. If the user searches for any of these options of words, the page links are made available to the user, Page Rank wise.

The most relevant search results' links will appear (Page Rank Wise). The link's title gives the idea about the information on

that particular page. On clicking the relevant link, user can directly access the desired information. After clicking on the search button, the link appearing at first position is at the rank 1st, the link at the second position is at the rank 2nd and so on.

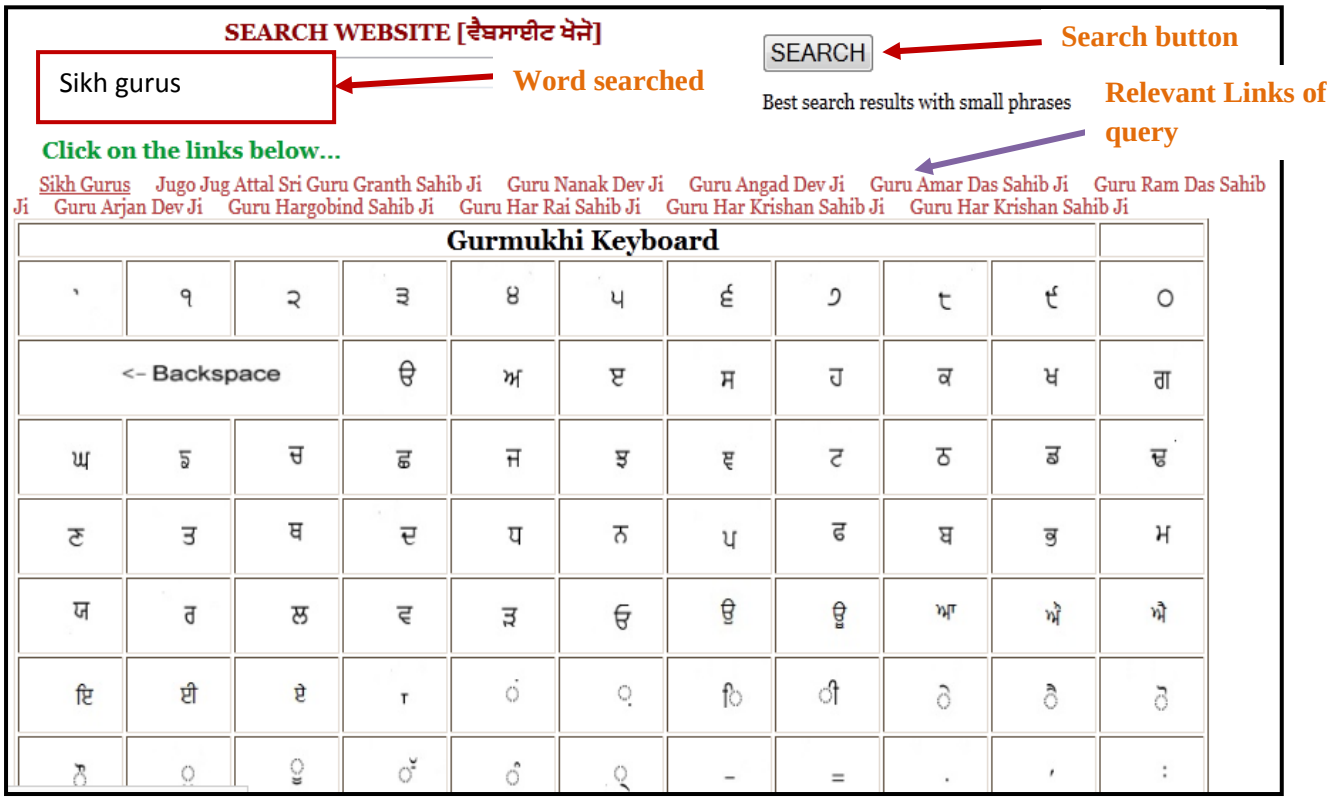


Fig 4: Search by Relevance (English) working

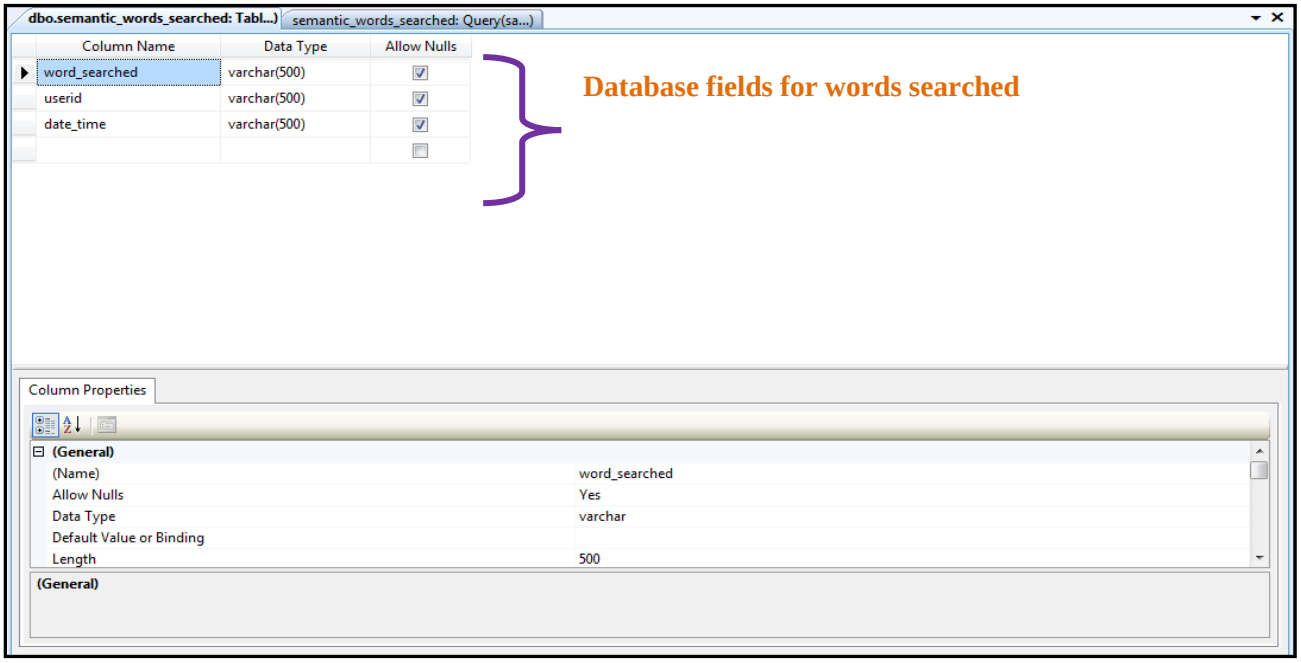


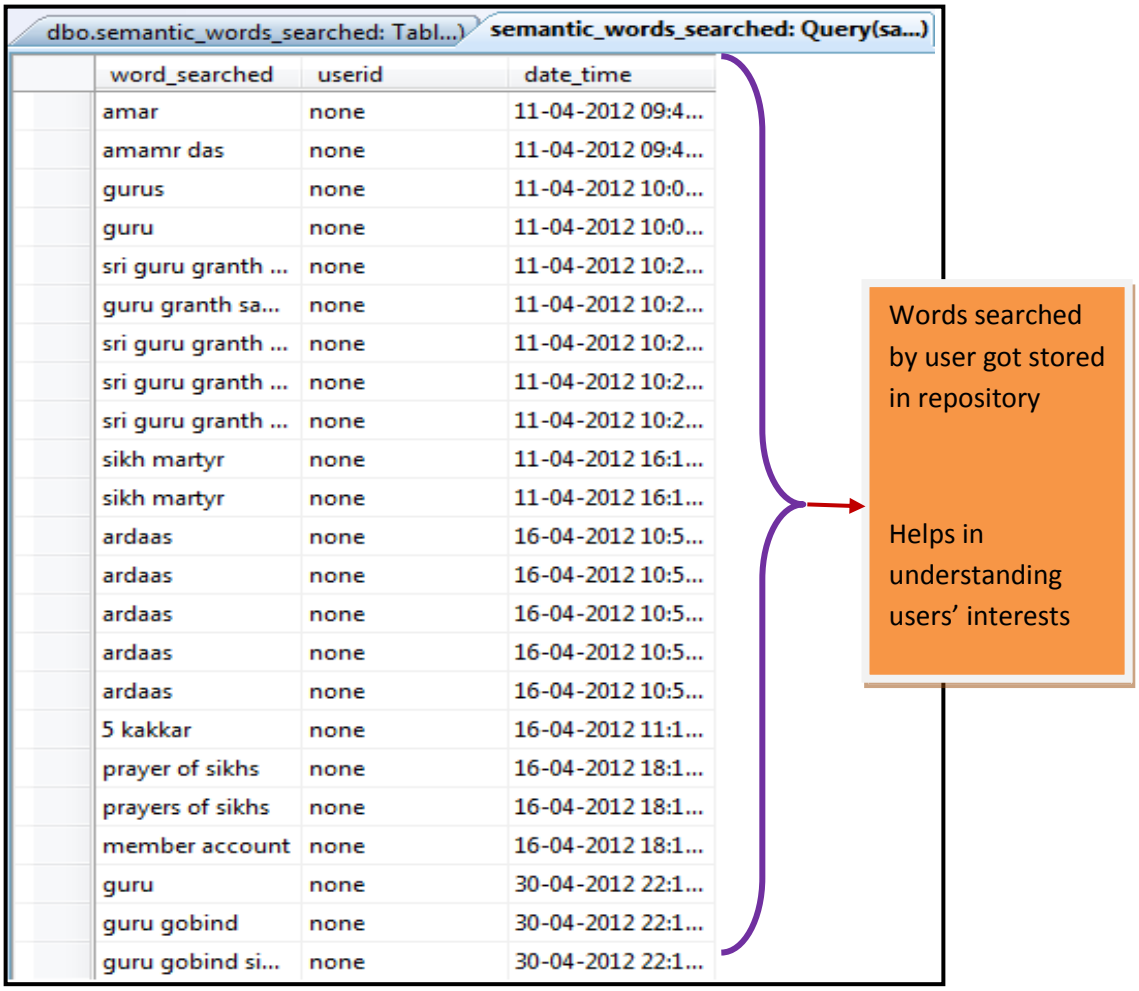
Fig 5: Semantics words searched (English) repository (Metadata)

Creating Repository of Searches by user

Users search cannot be determined completely beforehand. For this, the words searched by user as recorded in the repository. This helps sin knowing the user's interests much better.

This way, these words can be added to the options of words in semantics repository, by the administrator. This will make the search even better, every time the user searches for some information.

Next time, whenever the user searches for some information, we will be available with the user's search style already in advance. In this manner, user's search can be predicted to some extent beforehand.



word_searched	userid	date_time
amar	none	11-04-2012 09:4...
amamr das	none	11-04-2012 09:4...
gurus	none	11-04-2012 10:0...
guru	none	11-04-2012 10:0...
sri guru granth ...	none	11-04-2012 10:2...
guru granth sa...	none	11-04-2012 10:2...
sri guru granth ...	none	11-04-2012 10:2...
sri guru granth ...	none	11-04-2012 10:2...
sri guru granth ...	none	11-04-2012 10:2...
sikh martyr	none	11-04-2012 16:1...
sikh martyr	none	11-04-2012 16:1...
ardaas	none	16-04-2012 10:5...
ardaas	none	16-04-2012 10:5...
ardaas	none	16-04-2012 10:5...
ardaas	none	16-04-2012 10:5...
ardaas	none	16-04-2012 10:5...
5 kakkar	none	16-04-2012 11:1...
prayer of sikhs	none	16-04-2012 18:1...
prayers of sikhs	none	16-04-2012 18:1...
member account	none	16-04-2012 18:1...
guru	none	30-04-2012 22:1...
guru gobind	none	30-04-2012 22:1...
guru gobind si...	none	30-04-2012 22:1...

Fig 6: Semantics words searched (English) by user (Repository)

4.2 Search by Relevance in Punjabi Language on Cultural Portal

The cultural portal developed is related to Sikhism. Punjabi is the main language of Sikhs. Keeping this in consideration, search facility is also provided in Punjabi. For this Gurumukhi Keyboard is provided on the portal. On clicking the buttons of the Gurumukhi letters, the character clicked by user gets typed in the textbox (besides the 'SEARCH' button).

A number of options of different phrases are used to produce desired search results for the user. After completing the phrase user clicks the search button. The page is redirected to the desired page.

The options of words written in different ways have been provided to system in the code behind the module. If the user searches for any of these options, the page will be redirected to the desired page.

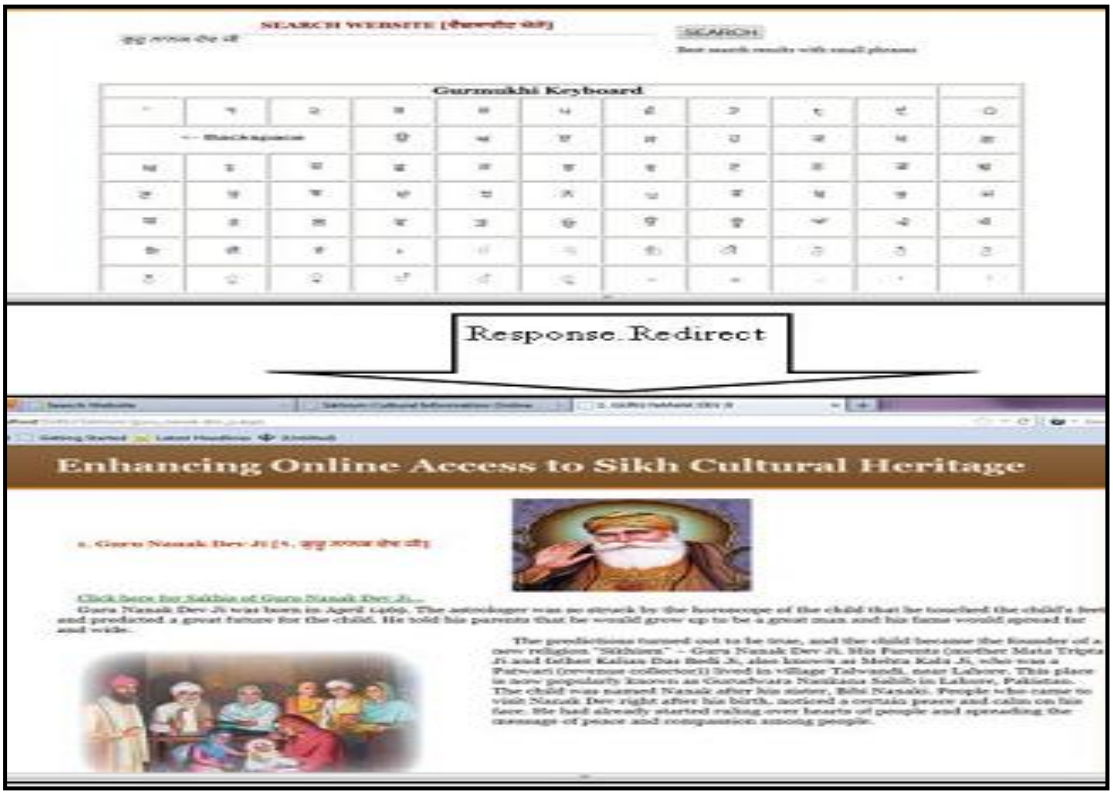


Fig 7: Search by Relevance (Punjabi) Implementation

Searching Metadata of tables used for Cultural Portal

If anyone wants to contribute some information about Sikhism to the cultural portal, he or she can request to search metadata of the tables containing such information.

If that information is not available on the cultural portal, the user can add that information, and content of the portal can be enhanced.

Along with searching metadata feature also contents of the tables are made available to the user on request. These contents enable the user to clearly identify that if the information available with him, is published on the portal or not, and contribute to the portal for enhancing information.

User searches for small phrases of word and the corresponding table's metadata and contents are made available to the user.

Figure 8 shows that search words can be stored against their corresponding tables, as the said information is stored in these corresponding tables.

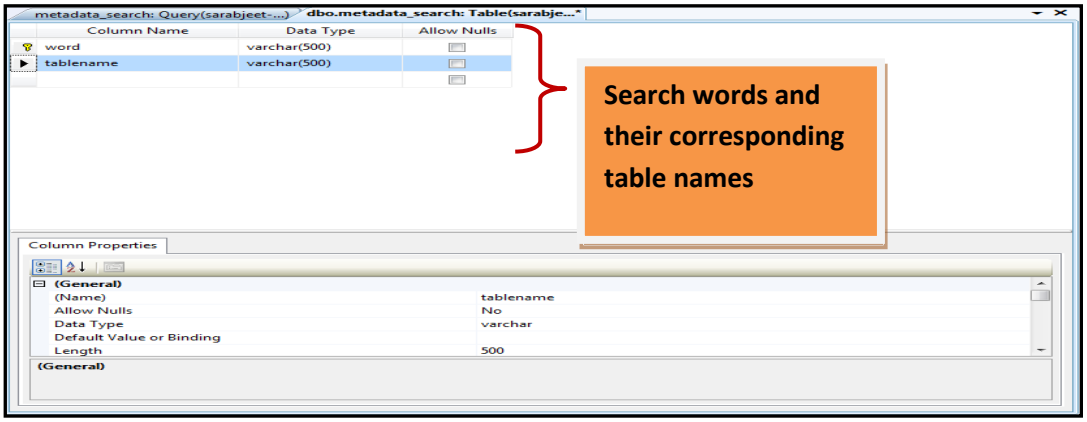


Fig 8: Searching Table Metadata

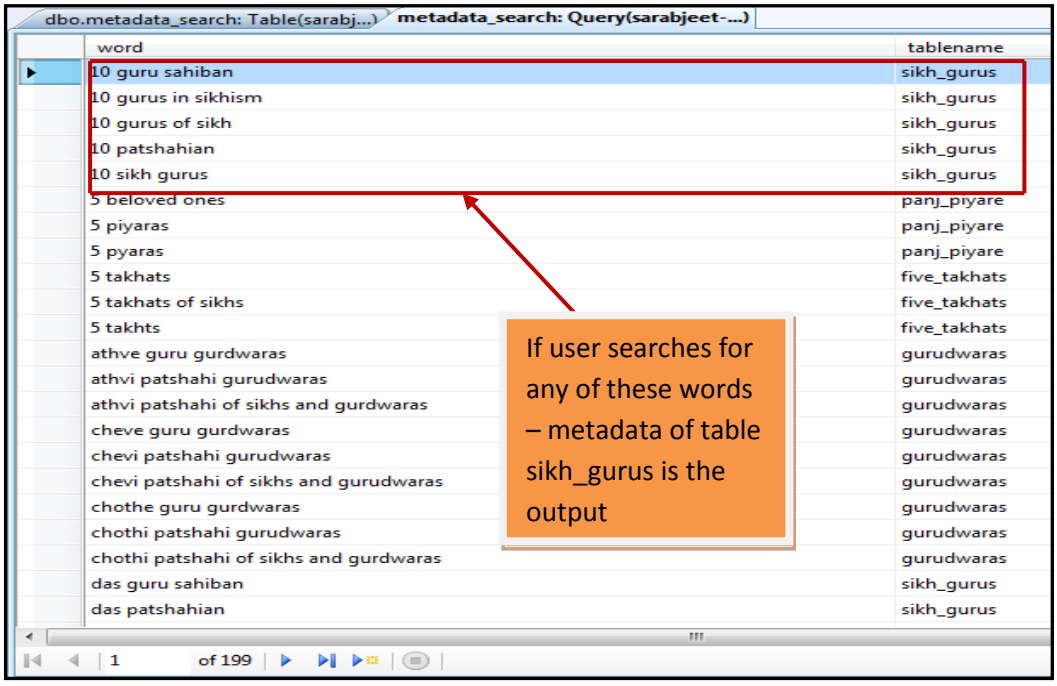


Fig 9: Searching Table Metadata (Repository)

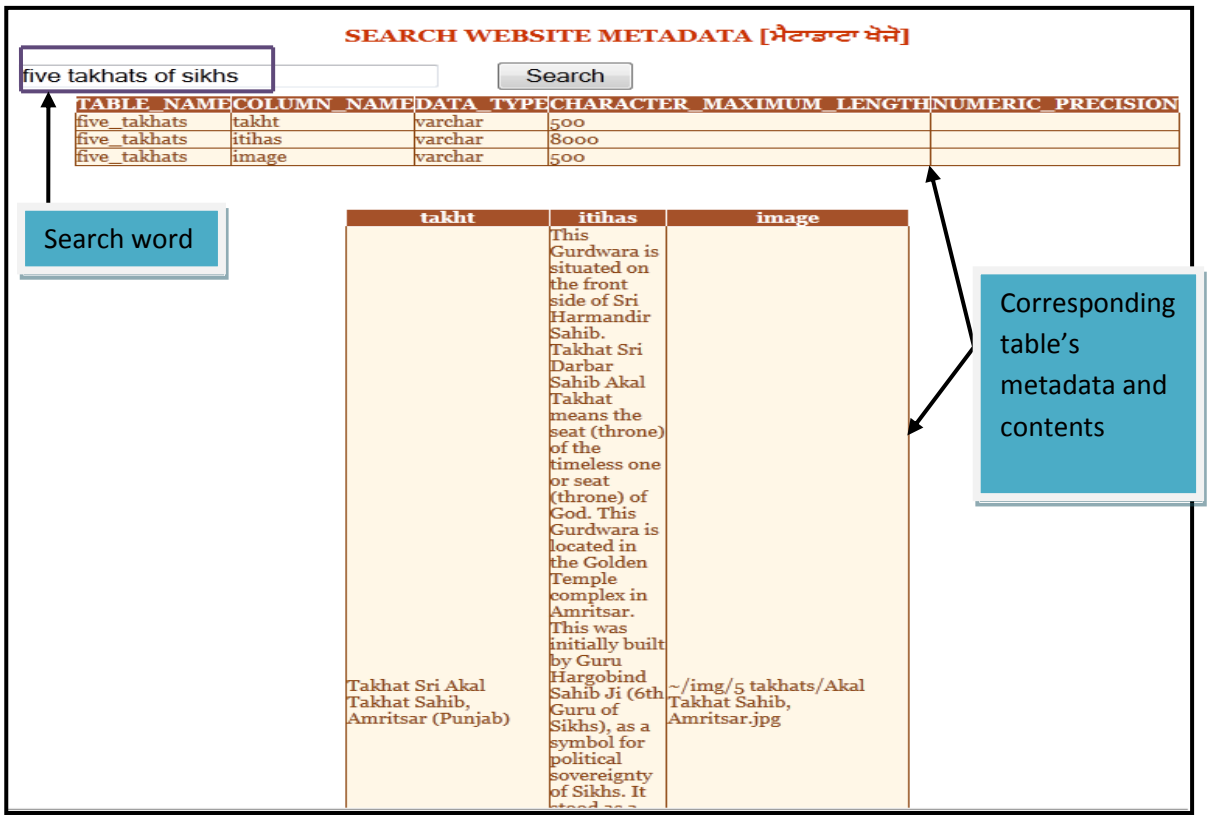


Fig 10: Searching Metadata of Tables for website information

Figure 9 describes the repository of search words of users and corresponding table names. These tables store the information about these corresponding search words. Figure 10 represents that if the user searches for a word; the metadata and contents of the table is made available to the user.

There is no end to knowledge. So, a provision to the user can be provided that, if the users want to contribute any new information to the cultural portal, they can do the same. But before contributing, they should know the existing information on the portal. For this they can search a word related to this kind of information; and get to know that whether the cultural portal has information about it or not. This way user can proceed further whether there is any need to add the information regarding the said cultural object or not.

5. CONCLUSION

WEB 3.0 supports the increased use of semantic technologies. [5] This paper aims at providing the search website feature that is offered in two kinds namely, Search by Relevance (Page Rank Wise) and Searching Metadata of tables. Search by Relevance feature provides the most relevant results (page rank wise) depending upon the words searched. This feature uses the concept of semantic relatedness (understanding the meaning of words and users need). Searching Metadata feature is also offered in this paper that helps the user know the exact cultural information available on the portal and contribute new information to this portal, which is not available on the portal while searching the table metadata and the complete data in the table itself.

6. REFERENCES

[1] Jagjit Singh, "The message of the Gurus – The Sikh Tree – Vol - 1". Published by B.Chattar Singh Jiwan Singh, 2006.

[2] Jorge Gracia and Eduardo Mena, "Web-Based Measure of Semantic Relatedness", IIS Department, Univ. of Zaragoza, Mar´ıa de Luna 1, 50018 Zaragoza, Spain.

[3] Nakatsu Ryohei, Tosa Naoko, A Computational Approach to Culture, cover story. Available at www.innovationmagazine.com

[4] Nova Spivak, Lew Tucker, Tim Boudreau, Henry Story, "Developing WEB 3.0", JavaOneSM Conference, 2007.

[5] Ora Lassilar, James Hendler," Embracing Web 3.0", Edited by Charles Petrie, IEEE Computer Society, 2007.

[6] Simone Paolo Ponzetto, Michael Strube, "Knowledge Derived FromWikipedia For Computing Semantic Relatedness", Journal of Artificial Intelligence Research 30 (2007).

[7] Ted Pedersen, Satanjeev Banerjee Siddharth Patwardhan, Maximizing Semantic Relatedness to Perform Word Sense Disambiguation. Available at http://static.msi.umn.edu/rreports/2005/25.pdf

[8] Kumar Singh Ashutosh, Kumar P Ravi, 'A Comparative Study of Page Ranking Algorithms for Information Retrieval', 2009.

[9] Sharma Kumar Dilip, 'Comparison of various web page ranking algorithms', UCSE, International journal on computer science and engineering, 2010.