

# 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'.

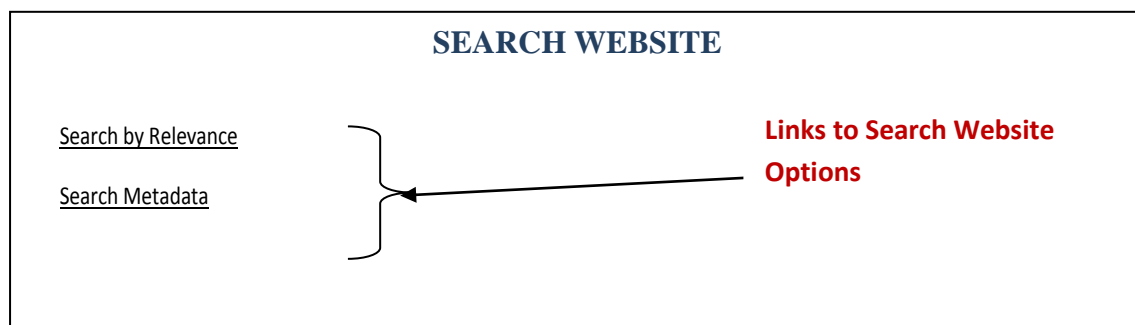


Fig 1: Search Portal Options

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.

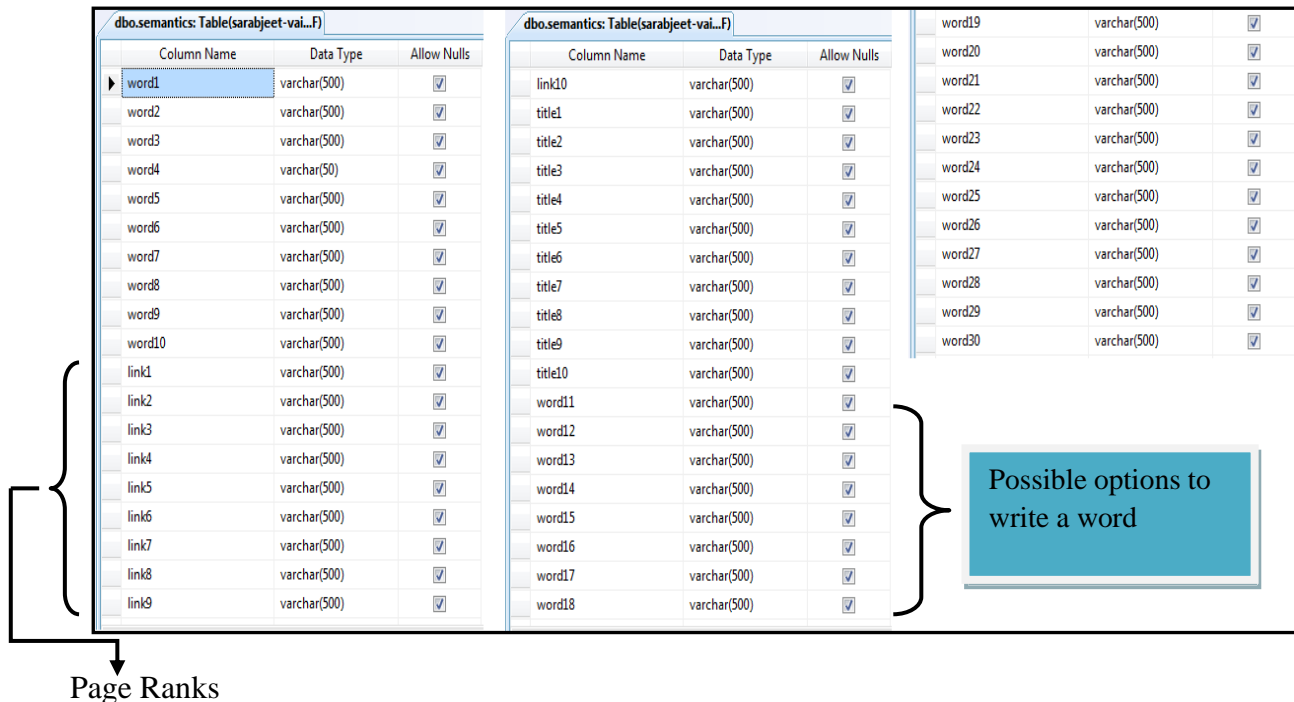


Fig 2: Semantics Repository Metadata

word1	word2	word3	word4	word5	word6	word7	word8	word9	word10	lii
guru	sikh gurus	10 gurus in sikh...	10 sikh gurus	ten gurus	ten sikh gurus	das guru sahiban	das patshahian	ten patshahian	10 patshahian	sik
5 pyaras	5 beloved ones	five beloved ones	5 piyaras	five pyara	five pyare	panj pyare	punj pyare	five pyaras	khalsa panth	pa
prayers	prayers	banis	sikh banis	sikh prayers	gurbani	sikh sangeet	hymns	holy hymns	paath	ba
guru nanak	guru nanak dev ji	pehli patshahi	pehli patshahi ...	nanak	patshahi pehli	founder of sikhi...	first guru of sikhs	1st guru of sikhs	NULL	sik
gurdwaras	gurudwaras	sikh shrines	sikh gurdwaras	historical sikh g...	historical gurd...	historical gurd...	historical sikh g...	sikh gurudwaras	NULL	gu
zaffarnama	letter of victory	zaffarnama	zaffarnamah	zaffar nama	zaffar namah	zafar nama	guru gobind si...	NULL	NULL	za
dasam pita	guru gobind si...	tenth guru of si...	bajan wale	dasmi patshahi	dasvin patshahi	dasvi patshahi	guru gobind si...	guru gobind si...	waho waho go...	gu
chali mukte	40 muktas	40 liberated ones	40 muktae	40 mukte	chaali mukte	mukte	40 muktee	chali muktee	NULL	40
karah parshad	kada prashad	karah prashad	kada prasad	karah prasad	sacred pudding	sacred pudding...	karah parsad	karah prasad	krah parshad	gu
guru angad dev ji	duji patshahi	dusri patshahi	doosri patshahi	bhai lehna ji	2nd nanak	angad	guru angad	second nanak	dooji patshahi	gu
guru amar das ...	guru amar das ji	guru amardas s...	guru amardas ji	third nanak	3rd nanak	amar das	amardas	teesri patshahi	tiji patshahi	gu
guru ram das s...	guru ramdas sa...	guru ram das ji	guru ramdas ji	fourth nanak	4th nanak	chothi patshahi	chauthi patshahi	bhai jetha	bhai jetha ji	gu
guru arjan dev ji	guru arjan	arjan	fifth nanak	5th nanak	5th patshahi	panjvi patshahi	panjvin patshahi	fifth guru of sik...	5th guru of sikhs	sik
sikh martyrs	sikh shaheed	shaheed	martyrdom in s...	NULL	NULL	NULL	NULL	NULL	NULL	sik
guru hargobind...	guru har gobin...	guru hargobind ji	guru har gobin...	6th nanak	sixth nanak	miri piri de maa...	miri piri de malak	sixth guru of sik...	patshahi chevin	gu
5 takhts	5 takhts	five takhts	five takhts	five takhtas	takhts of sikhs	sikh takhts	panj takhat	punj takhat	five takhts of s...	5
anand karaj	sikh marriage	marriage cerem...	marriage cerem...	sikh marriage c...	anand kaaraj	anand kaaraj ce...	laavan	4 laavan	four laavan in si...	an
ardas	sikh supplication	sikh invocation	sikh ardaas	sikh prayer	prayer of sikhs	ardas	sikh ardas	NULL	NULL	an
guru har rai sah...	7th nanak	seventh nanak	seventh guru of...	7th guru of sikhs	seventh patshahi	7th patshahi	guru har rai ji	har rai	patshahi satvi	gu
guru har krisha...	guru harkrishan...	8th nanak	eighth nanak	8th guru of sikhs	eighth guru of ...	guru harkrishan ji	eighth patshahi	harkrishan	har krishan	gu
guru teg bahad...	guru teg bahad...	teg bahadur	9th nanak	ninth nanak	9th guru of sikhs	ninth guru of si...	ninth patshahi	9th patshahi	patshahi ninth	gu
four sahibzade	4 sahibzade	guru gobind si...	dasvin patshahi...	dasvin patshahi...	guru sahib de s...	4 princes of gur...	four princes of ...	sons of guru go...	sons of guru go...	

Different ways to write a word

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.

**SEARCH WEBSITE [ਵੈਬਸਾਈਟ ਖੋਜੋ]**

Sikh gurus

**Word searched**

Click on the links below...

**Search button**

Best search results with small phrases

**Relevant Links of query**

[Sikh Gurus](#)   [Jugo Jug Attal Sri Guru Granth Sahib Ji](#)   [Guru Nanak Dev Ji](#)   [Guru Angad Dev Ji](#)   [Guru Amar Das Sahib Ji](#)   [Guru Ram Das Sahib Ji](#)  
[Guru Arjan Dev Ji](#)   [Guru Hargobind Sahib Ji](#)   [Guru Har Rai Sahib Ji](#)   [Guru Har Krishan Sahib Ji](#)   [Guru Har Krishan Sahib Ji](#)

,	ੴ	੨	੩	੪	੫	੬	੭	੮	੯	੦
<- Backspace			ੳ	ਅ	ੲ	ਸ	ਹ	ਕ	ਖ	ਗ
ਘ	ਙ	ਚ	ਛ	ਜ	ਝ	ਞ	ਟ	ਠ	ਡ	ਢ
ਣ	ਤ	ਥ	ਦ	ਧ	ਨ	ਪ	ਫ	ਬ	ਭ	ਮ
ਯ	ਰ	ਲ	ਵ	ੜ	ੳ	ੳ	ੳ	ਆ	ਐ	ਐ
ਇ	ਈ	ਏ	ੲ	ਂ	੍	ਿ	ੀ	ੇ	ੈ	ੋ
ੈ	ੳ	ੳ	ੳ	ੳ	ੳ	-	=	.	,	:

Fig 4: Search by Relevance (English) working

dbo.semantic\_words\_searched: Tabl...   semantic\_words\_searched: Query(sa...)

Column Name	Data Type	Allow Nulls
word_searched	varchar(500)	<input checked="" type="checkbox"/>
userid	varchar(500)	<input checked="" type="checkbox"/>
date_time	varchar(500)	<input checked="" type="checkbox"/>
		<input type="checkbox"/>

**Database fields for words searched**

**Column Properties**

<b>(General)</b>	
(Name)	word_searched
Allow Nulls	Yes
Data Type	varchar
Default Value or Binding	
Length	500
<b>(General)</b>	

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 in 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.

dbo.semantic_words_searched: Tabl...		semantic_words_searched: Query(sa...	
	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...

Words searched by user got stored in repository

Helps in understanding users' interests

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.

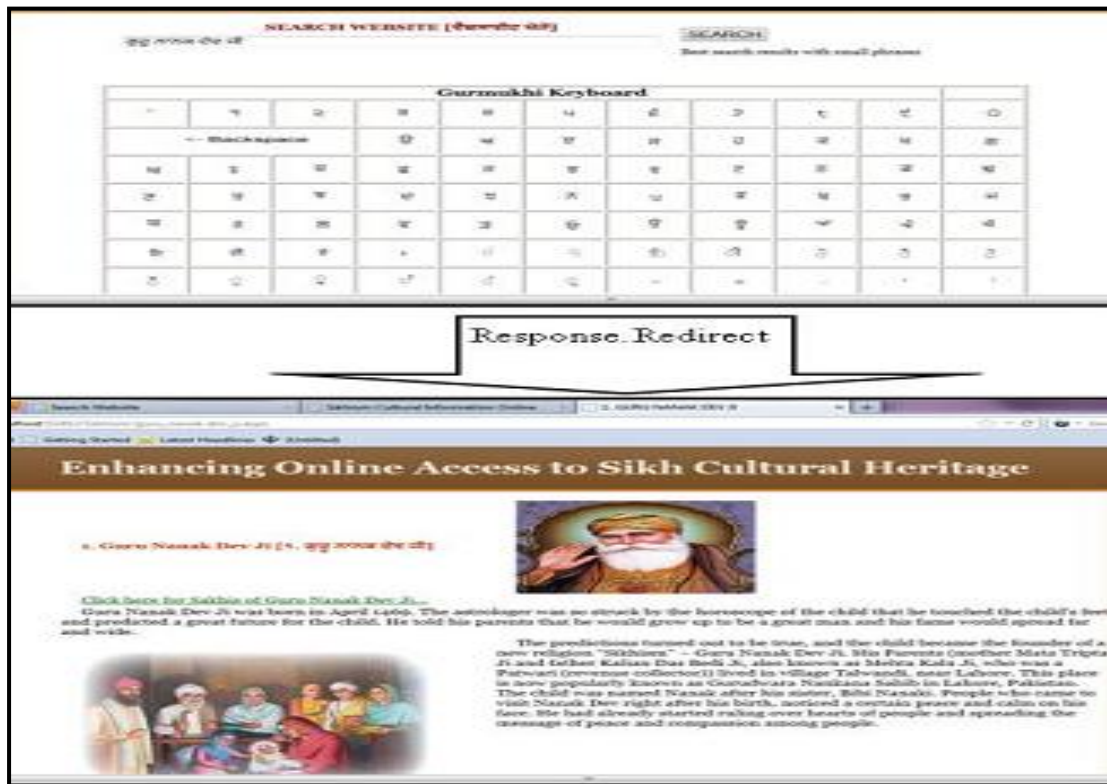


Fig 7: Search by Relevance (Punjabi) Implementation

## 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.

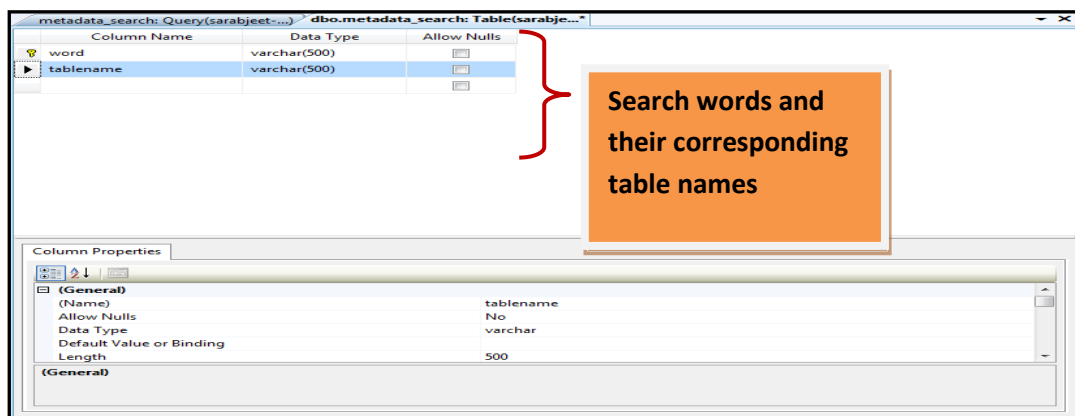


Fig 8: Searching Table Metadata

word	tablename
10 guru sahiban	sikh_gurus
10 gurus in sikhism	sikh_gurus
10 gurus of sikh	sikh_gurus
10 patshahian	sikh_gurus
10 sikh gurus	sikh_gurus
5 beloved ones	panj_piyare
5 piyaras	panj_piyare
5 pyaras	panj_piyare
5 takhats	five_takhats
5 takhats of sikhs	five_takhats
5 takhts	five_takhats
athve guru gurdwaras	gurdwaras
athvi patshahi gurdwaras	gurdwaras
athvi patshahi of sikhs and gurdwaras	gurdwaras
cheve guru gurdwaras	gurdwaras
chevi patshahi gurdwaras	gurdwaras
chevi patshahi of sikhs and gurdwaras	gurdwaras
chothe guru gurdwaras	gurdwaras
chothi patshahi gurdwaras	gurdwaras
chothi patshahi of sikhs and gurdwaras	gurdwaras
das guru sahiban	sikh_gurus
das patshahian	sikh_gurus

Fig 9: Searching Table Metadata (Repository)

**SEARCH WEBSITE METADATA [ਮੈਟਾਡਾ ਖੋਜੋ]**

five takhats of sikhs

TABLE_NAME	COLUMN_NAME	DATA	TYPE	CHARACTER_MAXIMUM_LENGTH	NUMERIC_PRECISION
five_takhats	takht		varchar	500	
five_takhats	itihas		varchar	8000	
five_takhats	image		varchar	500	

**Search word**

takht	itihas	image
Takhat Sri Akal Takhat Sahib, Amritsar (Punjab)	This Gurdwara is situated on the front side of Sri Harmandir Sahib. Takhat Sri Darbar Sahib Akal Takhat means the seat (throne) of the timeless one or seat (throne) of God. This Gurdwara is located in the Golden Temple complex in Amritsar. This was initially built by Guru Hargobind Sahib Ji (6th Guru of Sikhs), as a symbol for political sovereignty of Sikhs. It	~/img/5 takhats/Akal Takhat Sahib, Amritsar.jpg

**Corresponding table's metadata and contents**

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'ia de Luna 1, 50018 Zaragoza, Spain.
- [3] Nakatsu Ryohei, Tosa Naoko, A Computational Approach to Culture, cover story. Available at [www.innovationmagazine.com](http://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.