International Journal of Web Information Systems

Relevance status value model of Index Islamicus on Islamic History and Civilizations

Roslina Othman Ashraf Ali Salahuddin

Received 24 June 2014
Revised 11 October 2014
Accepted 12 October 2014

Relevance status value model of Index Islamicus on Islamic History and Civilizations

Roslina Othman and Ashraf Ali Salahuddin
Department of Library and Information Science, International Islamic University Malaysia (IIUM), Kuala Lumpur, Malaysia

Abstract

Purpose – The purposes of this study were to measure the relevance status of Index Islamicus, evaluate the semantic correlation between a query and documents and inquire the basis of its rank. Sorting the retrieved results from the most relevant to the least relevant is the common option of an information retrieval system. This sorting mechanism or relevance judgment is computed by measuring closeness of query with its documents.

Design/methodology/approach – Forming up 100 queries on Islamic History and Civilizations, with two indexing elements (keyword and concept), a laboratory experiment was generated on its first ten items of the rank. Throughout an experimental research design, the relevance status value formula was used to measure system-computed rank and compare it with mean average precision.

Findings – The results showed that the average status value of Index Islamicus's ranking on relevance criterion was 18 per cent effective in terms of retrieving precise documents. Despite the main focus of this study being only on one subject domain and the items calculated were only 1,000, this small percentage of its ranking mechanism proved that semantic correlations between queries with subject domain did not achieve the satisfactory level.

Research limitations/implications – Implication of this study could be a guideline for further research on ranking mechanism of other search engines because the limitation of this study was Index Islamicus being the only database, which was the focus of this study.

Practical implications – Throughout this study, Index Islamicus would be benefited knowing the status of its ranking mechanism as well as other databases can make further research on their own ranking method following this study.

Social implications – Researchers and vendors of online databases can ensure their users a true platform of search engine with a proper ranking list.

Originality/value – Relevance status value model for Index Islamicus on Islamic History and Civilization that allows the system to rank documents according to the match between document and query and gives the idea of a better index. The model improves the system's ranking mechanism, and promotes the use of semantic relationships. This research promotes the computation of relevance status value by domain for capturing subject-specific relevance criteria and semantic relationships.

Keywords Information retrieval, Index Islamicus, Islamic history, MAP, Relevance status value, RSV

Paper type Research paper

International Journal of Web Information Systems
Vol. 11 No. 1, 2015
pp. 54-86

1744-0084
DOI 10.1108/IJWIS-06-2014-0024

1. Introduction

Information storage and retrieval refers to the preservation of information and provision of ways to access it. When information and publications are stored online, it is important that keywords or indexing terms are assigned to the stored documents so that users can

easily find what they need (Abdulahhad *et al.*, 2012a, 2012b; Chu, 2010; Ouchetto *et al.*, 2012). Kettani and Newby (2010, p. 1) stated that:

> IR systems – is to provide a ranked list of the most relevant documents. This list is created based on matching the human expression of information need – the query – to a set of documents.

This means that, when a query or a search term matches with a document stored in the database, that document is retrieved. The ranking list is set of most relevant to the query to the least relevant. All these matching activities and ranking orders are prepared by the system computed mechanism. If system computed ranking does not contain true relevance criteria where most shared elements of a query and documents are the most related, the user would be led to another title or document that he does not want really, or user's expected title would appear at the bottom of the ranking list where it would escape his attention.

Therefore, after the indexing process is done, there must be an evaluation to test the effectiveness of the platform based on the level of suitability between the documents' titles and the queries terms or vice versa (Meadow *et al.*, 2007). This study would like to perform an evaluation on *Index Islamicus* to test its default ranking model through relevance status value (RSV) method that would examine the system's relevance to its query and documents' titles. Such an evaluation has been generated on various databases or Web searches in many previous researches. Examples of these are Cross-Language Evaluation Forum (CLEF) test by Abdulahhad *et al.* (2011, 2012a, 2012b) through RSV; relevance ranking for e-government services retrieval or the work by Ouchetto *et al.* (2012); the study of relevance ranking of video comments on YouTube by Serbanoiu and Rebedea (2013); and also the research on document ranking through enterprise search (ES) by Chunchen and Jianqiang (2012).

Despite all the studies and evaluations, the process of information retrieval (IR) and relevance judgment is always uncertain and subjective (Abdulahhad *et al.*, 2012a, 2012b; Manning *et al.*, 2009). For example, person *A* might see the relevance of his query "Sharia law" with the title of "Sharia and national law in Saudi Arabia", but to person *B* this title may not be relevant because he needs materials on the study of Sharia law in general. Figure 1 shows how *A* considers his relevance search, and Figure 2 shows how *B* considers his relevance search by the same query.

In this case, information retrieval process is uncertain. It cannot satisfy everyone and always depends on user's judgment (Ismail *et al.*, 2000). Nevertheless, according to Abdulahhad *et al.* (2012a, 2012b, p. 3), "to estimate the certainty of an implication and to offer a ranking mechanism, another component should be added" and that is the method of relevance status of the default rank by estimating the closeness of query and document. In another view point, there should be a method to check on the ranking mechanism of information retrieval system (IRS) (Meadow *et al.*, 2007) to ensure its suitable performance.

However, as IRS always focuses on two sets of document: the relevance to query and the irrelevance documents to query (Abdulahhad *et al.*, 2012a, 2012b; Manning *et al.*, 2009), and the suitability of the foundation of all IR models is "matching" (Chu, 2010), this research's main target is to evaluate the matching model according to retrieval ranking of *Index Islamicus* (an online database) through RSV formula of Abdulahhad *et al.* (2012a, 2012b). Accordingly, each information retrieval platform gives the

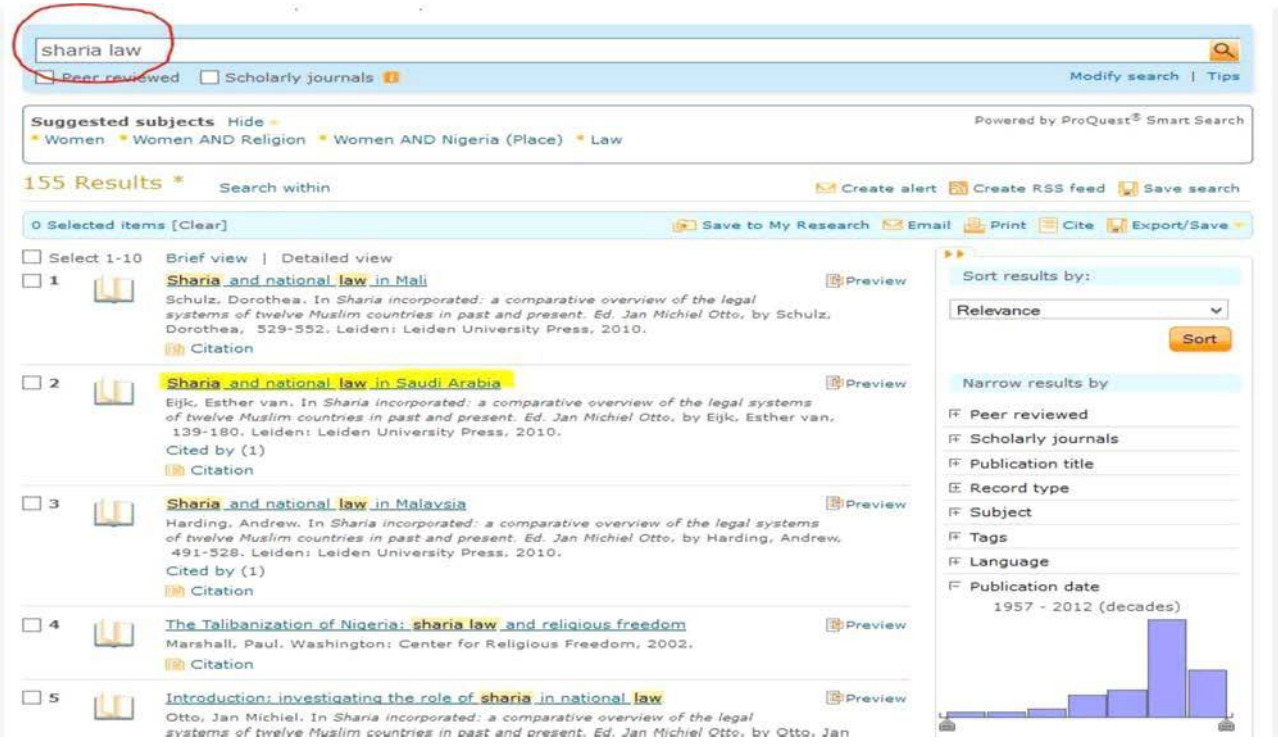


Figure 1.
Relevance judgment between two users

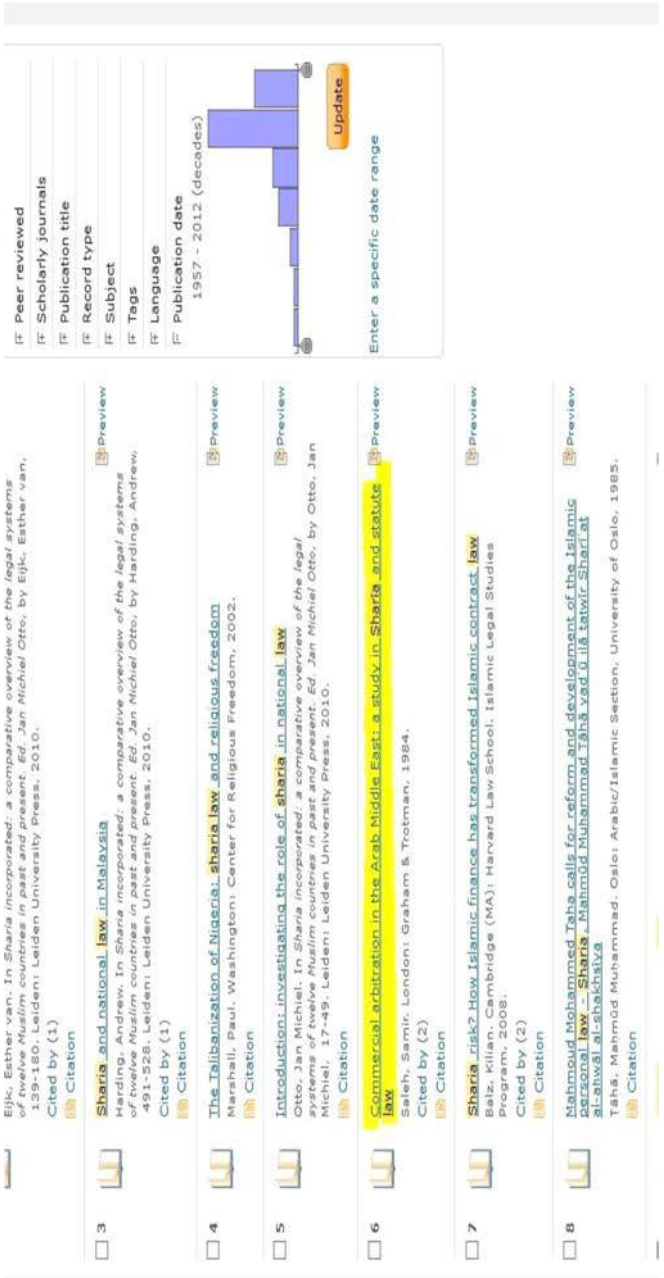


Figure 2.
Relevance judgment between two users

opportunity to store data and let the user come up with their own query search term to find it and, when the user is able to perform a search through standard form of query, all the purposes of an IR is to find out the "match" of that query with its existing documents and to give a comparative rank to the retrieved results (Chu, 2010).

Thus, this study would research on one of the popular online databases named "Index Islamicus". It sets to find out the semantic relation of "Index Islamicus" with natural query on a specific subject domain – *Islamic History and Civilizations*. Having searched on the system with some experimented queries, the matched or related titles would be justified throughout RSV. Then, it would find out reasonable ranking of those retrieved results which are most "semantically related" to the queries and comparatively show the RSV of *history and civilizations* subject domain in *Index Islamicus*.

1.1 Index Islamicus

Index Islamicus is a bibliography database of publications about Islam and the Muslim world. It covers almost 100 years of publications on the world of Islam. Over 3,000 journals are monitored for inclusion in the database, together with conference proceedings, monographs, multi-authored works and book reviews. Journals and books are indexed down to the article and chapter level. Over 462,012 records exist in *Index Islamicus*, according to the July 2013 update. The publications have been covered since 1906, and it updates the entire database quarterly (Ahmad, 1981; Anwar, 2001; CSA Illumina, 2013). The research is specifically done on the subject area of *Islamic History and Civilizations*. In this domain, around 9,393 titles were found from the entire collection of 462,012. When it was searched with *truncation character* ("Islam* history"), the total retrieved results were 13,002. So, the approximate collection of Islamic History and Civilization in *Index Islamicus* is about 9,000 to 13,000 entries. This research would formulate some queries regarding this domain and search on *Index Islamicus*.

1.2 Relevance status value

RSV is a method to evaluate the systems' ranking according to its relevance between documents and users' query. Through this method, the retrieval perfection of a database can be judged to find better index terms. It is a method where relationship between a document and a query is measured. The ranking value of RSV is considered the ranking of a document according to query. For example, the most similar document with its query is the first document in retrieving list as the most relevant. As queries are formed with two indexing elements for this research, the RSV ranking is the actual evaluation to justify the relevance status of this database (Index Islamicus). RSV shows how a document could be semantically closer to its query (Meadow *et al.*, 2007; Abdulahhad *et al.*, 2012a, 2012b). In RSV, documents are symbolized as "d" and queries as "q". So, the formula of RSV starts with $d \cap q$, where the most shared elements in documents and queries are the base of RSV. The result of RSV determines whether the system should retrieve the document or not and ranks the retrieval results according to the highest match among all others throughout its relevance with queries (Abdulahhad *et al.*, 2011; Meadow *et al.*, 2007). Hence, the *Index Islamicus* is evaluated on the subject of *Islamic History and Civilizations*. The database would retrieve the most related results to the queries and put them in rank. The evaluation of this relation or rank would be justified by RSV, and then re-ranked the retrieved items through its formula to measure the semantic relation and show the average precision (AP).

2. Significance of this study

In the IRS, the chances of a better match between documents and queries depend on its exhaustivity and specificity (Olson and Boll, 2001). In information representation and retrieval (IRR), exhaustivity indicates when subject term covers broad extent of semantic correlation; which means, the recall of search result would be higher, but when the breadth of subject matter is absent, it means that exhaustivity is low and hence the result would be lower. Accordingly, "specificity" refers to a situation in IRR when the subject term is specific and focused, such as when the focused term of a document is missing from its concept, the query cannot be counted as relevant and, hence, the recall may be higher, but precision will not be acceptable as expected (Abdulahhad *et al.*, 2012a, 2012b; Blanke and Lalmas, 2011; Kazai and Lalmas, 2006; Xiaohui *et al.*, 2010). To make the search effective, we need matched and represented information in the system with the query submitted by the user (Chu, 2010). According to Meadow *et al.* (2007), there should be a mechanism that can evaluate the matching relation between a user query and the document. One of the evaluative mechanisms is the RSV, which allows the system to rank documents according to the match between the document and query and gives an idea of a better index. Through this mechanism, users have a chance to evaluate the system's ranking criteria by comparing the ranking of RSV which contributes to the method emphasizing the words repetitive occurring in the text, title and index.

As history is heavily related with culture and heritage, the storage and retrieval process of this body of heritage on online database should consider the same factor. Across the globe, such as in the USA, the European countries and the Middle East, research and projects on culture and heritage have gained priorities. Hence, semantically correlation of user's judgment (query) with the collection of *Islamic History and Civilizations* is a necessary thing to work on for the purpose of better retrieval. Throughout this study, *Index Islamicus* would be benefited knowing the status of its ranking mechanism as well as other databases can make further research on their own ranking method. In this case, this effort can be considered a big impact on online database's search engine and effectiveness of ranking mechanism.

3. Problem statement

The correlation of documents and queries is a significant point of an online database in serving its stored data to the users. In other words, when a document shares most of its elements with the query, the chances of that document to be in ordered or ranked results is high. Similarly, when a query is focused and more specific on a subject theme, the ranking list of retrieved documents for that query would be more accurate (Abdulahhad *et al.*, 2012a, 2012b; Chu, 2010). To comply with this requirement for a better relevance ranking, a database or IRS should contain highly specific index terms (Ismail *et al.*, 2000) or thesaurus (Feldvari, 2005; Sato, 2007). However, *Index Islamicus* is one of the oldest databases which started in 1958 covering the publications from 1906 and uses some controlled index terms in Islamic domain through ProQuest after the Library of Congress Subject Heading (LCSH) (Salim *et al.*, 2010). The ranking mechanism of retrieved result of this database works in two ways:

(1) Through ProQuest four thesaurus [ERIC, MeSH, LISA and ProQuest (subject)].

(2) Through Interaction with its index terms and citations (as *Index Islamicus* does not provide full text as well as abstract).

As none of its thesaurus (out of four) specialized on Islamic domain particularly and match-making of query happens through its citations instead of full-text document or abstract, the relevance ranking of this database may mislead the user when highly shared elements of a query match with much more words-containing document (full text or abstract) rather than matching with a title and index terms (citations) only.

In addition, the evaluation of *Index Islamicus* ranking and retrieving mechanism has not been checked except for its index term descriptors (Kopycki, 2003). Therefore, this study performs an evaluation on its ranking mechanism through RSV method to measure its semantic relations with the queries and citations. The focus of this evaluation is the specific subject domain, which is the *Islamic History and Civilizations*.

3.1 Research objectives

- To determine RSV for *Index Islamicus* on the subject of *Islamic History and Civilizations*.
- To examine the semantic relationship between documents and queries on the subject of *Islamic History and Civilizations*.

3.2 Research question

Objective 1: "To determine RSV for Index Islamicus on the subject of Islamic History and Civilizations":

- What is the relevance judgment criteria commonly applied to the subject of *Islamic History and Civilizations* in *Index Islamicus*?
- What are the available controlled vocabulary tools or thesaurus for *Islamic History and Civilizations* of *Index Islamicus*?

Objective 2: "To examine the semantic relationship between documents and queries for the subject *of Islamic History and Civilizations* in *Index Islamicus*":

- What is the RSV for *Islamic History and Civilizations* in *Index Islamicus*?
- What are the differences between RSV ranking and *Index Islamicus* system's default ranking?

3.3 Assumptions

This research is conducted based on the hypothesis of the problem statement and the objective of the research where it is said that the more shared elements or concepts a document and a query have, the more related and closer they are (Chu, 2010; Abdulahhad *et al.*, 2011). In other words, the much more elements of a d (document) contain the q (query), the more closely or semantically related they are (d, q) and that would result in the higher d in retrieval ranking. Kettani and Newby (2010, p. 2) "The closer a document to a query is, the higher its rank". Accordingly, the less closer a document with the queries, the lower its relevance ranking and which would be useless for the users (Chawla and Bedi, 2008).

By taking this hypothesis into account, the RSV will evaluate the matching function and ranking criteria of *Index Islamicus* with the query on the subject of *Islamic History and Civilizations*.

4. Literature review

4.1 Background

Online databases are the widest place for information storage. *Index Islamicus*, one of the online databases, started its journey in 1958 and was built by the Middle Eastern librarians and bibliographers. In early twenty-first century it started a transformation from CD-ROM platform into Web and that was the beginning of the relevance search and retrieval issues (Ahmad, 1981; Anwar, 2001; Kopycki, 2003). Up to today, there are several criticisms made by few researchers on its options and modules rather than on the entire relevance testing or query judgments. Hence, this research proposes an evaluation criteria, the RSV (Abdulahhad *et al.*, 2011) to measure *Index Islamicus* semantic relation with real query. Subsequently the collection of *Index Islamicus* has grown, and today, it has reached more than 462,012 entries (CSA Illumina, 2013).

The literature review addresses three areas of research on the relationship between the semantic relation of queries and documents and the system evaluation. In the first section, research related to the early studies on IRS and its progress is discussed. In the second section, there are discussions on RSV, popular IRS evaluation models and ranking methods. The final section explains the effective methods for this research strategy and vocabulary controlled tools on *Islamic History and Civilizations* of the *Index Islamicus*.

4.2 Early studies on IRS

According to Chowdhury (2010), the first IRS evaluation was made in 1953. Until 2004, several studies have been conducted on the evaluation of indexing systems. Later on, Cranfield Projects and the MEDLARS study came through with a big contribution on the retrieval evaluation such as recall, precision and inverse relationship.

However, because indexing is a practice of creating comprehensive entry points for information (Ansari, 2005), after experimenting on the relevance of index term specificity, Kim (2006, p. 1227) stated in his result that, "documents should be indexed with specific terms in order to be retrieved as relevant documents", where 64 per cent cases in his research were correctly classified because of following specificity, which is narrowing down to specific subject term.

Huang and Zhang (2010) proposed a semantic tree model for queries in a way that was assumed better workout in recall and precision. In their method, they built a tree of semantic keywords based on WordNet, where they gave each keyword an expansion synonym and set a weight on it so that it strengthens the users' intent to pick up the best query for search such as Tree of Associational Semantics Model (TASM). Their result showed that this combined query expansion is better than the *tfidf* method.

In addition, Chatvichienchai and Tanaka (2009) came up with another search technique where their queries were built according to document type, search terms and their semantic relationship. In their research, they experimented on Microsoft office applications and they built an XML-based search indexes which collected the search terms from office applications and found their semantic relationship to generate an effective search to locate the specific document.

There is a similarity of these researches in the semantic relationship of a document and its queries. All those researches were done to improve the precision rather than just focusing on recall. Many researchers agreed that traditional search tools are giving opportunity to have higher recall but with comparably low precision.

On this assumption, Huang and Zhang (2010, p. 1) stated that:

> Traditional methods for query expansion often get into a great trouble improving both the recall and precision. Along with the increase of the size of the query set, the recall shows rapid progress while the precision usually drops down in a way.

Similarly, Chatvichienchai and Tanaka's (2009, p. 1) asserted:

> Traditional search tools based on conventional IR techniques (employing keyword and phrase matching between the query and index alone) tend to offer high recall and low precision. The search users are faced with too many irrelevant results.

In both of these experiments, recall is not all a user wants; the matter of search result and expectation is the precision. Likewise, precision does not happen because of expanded query in Huang and Zhang's view. Besides, irrelevant results have also proven to bother user due to the query matching with index term in Chatvichienchai and Tanaka's work.

On the other hand, it is said that IR is an uncertain process; it is imperfect to represent user's needs, and the relevance judgment always depends on external factors, which is the user knowledge and user choice of picking query (Manning *et al.*, 2009; Abdulahhad *et al.*, 2012a, 2012b). But, to estimate the assessment and the accuracy of close perfection or the system's performance, mechanisms like vector space model, Boolean logic model, probability model, and so on, are implemented based on necessity (Chu, 2010).

4.3 Ranking mechanism of IRS and RSV

According to Chu (2010) and Abdulahhad *et al.* (2011), the more shared elements a document and a query have, the higher the probability of relation between that document and the query get and more semantically closer they are. In the authors' view, each document should satisfy the theme of query, and the theme of document should be related to the query as well. In other words, ranking or similarity of documents occurs by calculating the distance between the query and the document, and all the retrieved documents are ranked or ordered based on the distance with its query (Ansari, 2005; Chatvichienchai and Tanaka, 2009; Huang and Zhang, 2010; Kettani and Newby, 2010).

Chowdhury (2010) stated that there have been five major areas of evaluation done since 1953 on IRS, and most of those are centered on indexing language, search technique, output ranking, term weighing and cost effectiveness. Similarly, relevance measurements on recall, precision, novelty, serendipity, multiplicity, *tf-idf* and RSV have also been used comparatively. Among these, recall and precision are the very common and mostly used evaluation methods in IRS (Nasraoui and Zhuhadar, 2010).

The most widely used relevance-based measures are recall and precision, which were introduced by Cleverdon in his Cranfield test. Recall calculates the ratio of relevant items retrieved from the corpus, and precision calculates the ratio of relevant items retrieved in a search. Recall indicates the system's ability to retrieve relevant items, while precision indicates the ability to reject irrelevant items focusing on precise result. Novelty calculates relevant items and new items in a search where multiplicity calculates relevance items and useful items in a search, and serendipity measures the irrelevant items and useful items in a search (Chowdhury, 2010; Manning *et al.*, 2009; Othman, 2009).

All these methods are under the category of performance measurement in relevance judgment and recalling ability of a search system. The problem of these methods is that

it needs comparison to evaluate a system, such as according to User *A*, the relevant items are *X*, and according to User *B*, the relevant items are *Y*; hence, the calculations are proceeded based on how much relevant/irrelevant items are found and which items are relevant and irrelevant, useful or not useful according to Users *A* and *B* and so on. Therefore, this study ignored these methods and aims to evaluate using another approach which measures system's relevance and ranking through a pure mathematical formulation where no real users are needed.

Consequently, variant of *tf-idf* is a set of measurement to assign a weight for each term and each specific document (Manning *et al.*, 2009) so that it turns out to be an appropriate index term for a specific corpus for some specific documents. Indicating "specific" means, *tf-idf* is computed to develop keywords or term frequency for a specific document with document frequency on a specific database. For example, "Islamic law" is a term of a document *A*, this term is measured by document frequency on a database *F*. Hence, the *tf-idf* will come up with a weight for this term (Islamic law) for the document *A* of database *F*. This measurement put too much focus on finding the most weighted term for a certain document to search on a certain database that can only be used as query for that database only. In that query setting term or to set indexing term, Huang and Zhang's TASM is better than the *tf-idf* method (Huang and Zhang, 2010). But they all are about assigning terms or queries rather than to evaluate a system based on the relation between query and documents.

Thus, to achieve the research objectives, other relevance measurement standard like Boolean Logic Model, Vector Space Model and Probabilistic model was not performed either. Although these models are standardized for similarity measurement of relevance between queries and documents, they have many limitations unlike RSV, which has been discussed in details by Chu (2010) and in Kettani's and Newby, 2010 research. Furthermore, the RSV model has been applied to generate the ranking estimation against retrieved documents or all documents in a corpus. Using three types of indexing elements on *ad hoc* image-based retrieval task, and case-based retrieval task, the authors declared the effective and successful performance of this model (Abdulahhad *et al.*, 2011, 2012a, 2012b).

Consequently, researches have shown the importance of semantic relationship, matching model and the evaluation method after the first development of IRS in 1953 (Chowdhury, 2010). Through all the reviews, it is perceived that the IRS is important in aspects of storing information and documents as well as retrieving precise results. But no evaluation method was found to have been generated on any Islamic database's performance on retrieving Islamic works and publications like what was done on Microsoft office applications by Chatvichienchai and Tanaka (2009) to solve high recall and low precision problem in finding office document in large databases; experiments of Abdulahhad *et al.* (2011, 2012a, 2012b) by RSV method on a medical database which contains medical images through documents; examining match between assigned descriptors and medical theses by Ansari (2005) and Dinh's *et al.* (2013) research on effectiveness on biomedical indexing and retrieval based terminologies as well as researches on *Index Islamicus* like Kopycki (2003), Anwar (2001) and Salim *et al.* (2010). There were few researches done on *Index Islamicus* but they were not specifically on the same objectives as this research that has been aimed to measure RSV or assessment of ranking mechanism.

4.4 Analysis and conclusion

As a result, this research is going to perform an experiment on *Index Islamicus* for its RSV on the subject of *Islamic History and Civilizations*. The previous research and evaluations on IRS have proven that to get better precision on retrieval results and semantically related ranking, there should be a close relationship between documents and queries and to find out the comparison of system computed relevance rank with RSV rank, mean average precision (MAP) is very widely used and a stable evaluation measure to the researchers (Park, 2011; Yisong *et al.*, 2007).

However, *Index Islamicus* is a database where over 100 years of publications of Islam and the Muslim world with more than 3,000 journals are being monitored. There are more than 462,012 titles exist in *Index Islamicus*. It is necessary to take an effective evaluation measure on this database on *Islamic History and Civilization* even though it seems to have a very small impact in comparison to the entire database size and subject matters.

This research focuses on the RSV according to the hypothesis generated by Abdulahhad *et al.* (2011) to find out the relevance judgment criteria commonly applied to the subject of *Islamic History and Civilizations* in *Index Islamicus* using MAP comparison evaluation measure. The questions posed by this work are: what are the available controlled vocabulary tools or thesaurus for *Islamic History and Civilizations* of *Index Islamicus*? What is the RSV for *Islamic History and Civilizations* in *Index Islamicus*? What are the differences between RSV ranking and *Index Islamicus's* system-computed ranking?

The most widely used vocabulary controlled tools for Islamic History (or generally for Islam itself), is the LCSH. For further reference on Islamic ontology, besides LCSH, researchers usually also refer to *Index Islamicus*'s thesaurus (Salim *et al.*, 2010). Therefore, the research combines LCSH and *Index Islamicus*'s general thesaurus to generate queries for RSV evaluation.

5. Methodology

5.1 Research design

This study was an experimental research design where *Index Islamicus* was measured through a mathematical evaluation formula. The RSV is a model that was used and fifty titles were selected on the subject of *Islamic History and Civilizations* from *Index Islamicus* itself after a subject heading search "Islam* history" to form up queries. Two indexing elements were formed through those titles which are "keyword" and "concept" to use as query search. There were ten expected results calculated after each search. Then, RSV examined the search results on the basis of its mathematical formula to find out the relevance ranking of the database on this subject.

The entire experiment was done as a laboratory testing where no natural sampling or users were involved.

5.2 Research instruments

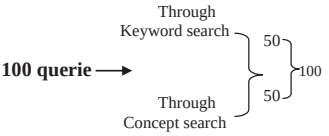
There were two basic instruments used. The first instrument is the chosen queries (keywords and concept), and the second instrument is the RSV formula itself with MAP measurement.

The chosen queries are the two indexing elements: keywords and concepts. The keywords were randomly formed according to the respective title's area and to improve subject searching, truncation was used because it is proven to be a better solution even

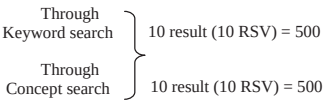
though sometimes it is not the best solution (Olson and Boll, 2001). It gives the opportunity to go through root word and helps to find near match, and it is also necessary to include stemming (Meadow *et al.*, 2007; Radovanović and Ivanović, 2008).

The second element of indexing is *concept*. The concept in this research was used as the controlled vocabulary. As concept is a form of notions or terms which is used independently to organize perception and knowledge (Abdulahhad *et al.*, 2011), it was not fair to justify an entire database (*Index Islamicus*) forming up through self-made concept query. Therefore, using 50 titles, both the "concept" search and the "keyword" search was formed comparing LCSH and *Index Islamicus* thesaurus, so that the queries used had a standard form to maintain disambiguation of proper subject term.

The total queries were 100. For each query, the search was made two ways through two indexing elements – "keyword" and "concept". So, for 50 queries, there were two searches; one, through "keyword" and another one, through "concept" which divided the **queries** into two (keyword searches + 50 concept searches). Later, having computed RSV for each keyword and concept, both were evaluated by MAP to compare them with system's default ranking.



For each "keyword" search, the first ten results were collected to compute seemingly the highest-ranking RSV according to *Index Islamicus* search outputs.



So, a total queries of 100 were divided into two sets. *Keyword set* and *Concept set*. Each query contained 20 titles through two searches (keyword + concept) (Figure 3).

To find the highest RSV for each individual query and compare it with system-computed ranking, both "keyword" and "concept" were evaluated by AP. Then, all 50 queries' AP was summed up to determine the difference with system's default ranking (Figure 3).

The mathematical formula of RSV to evaluate retrieved items through keywords is:

$$\mathrm{RSVk}\,(\mathrm{d},\mathrm{q}) = \|\mathrm{d}\cap\mathrm{q}\|\mathrm{k} \times \left(\sum \frac{\mathrm{N}}{\mathrm{Nk}} \times \frac{\mathrm{fd},\mathrm{k}}{\|\mathrm{d}\|} \times \|\mathrm{k}\|\right) \tag{1}$$

The formula through concepts is:

$$\mathrm{RSVc}\,(\mathrm{d},\mathrm{q}) = \|\mathrm{d}\cap\mathrm{q}\|\mathrm{c} \times \left(\sum \frac{\mathrm{N}}{\mathrm{Nc}} \times \frac{\mathrm{fd},\mathrm{c}}{\|\mathrm{d}\|}\right) \tag{2}$$

To evaluate further average over all queries through AP:

$$\mathbf{AP} = \frac{\mathrm{RSVi(k/c)} + \ldots \mathrm{RSVi(k/c)/k}}{\mathrm{k}} \tag{3}$$

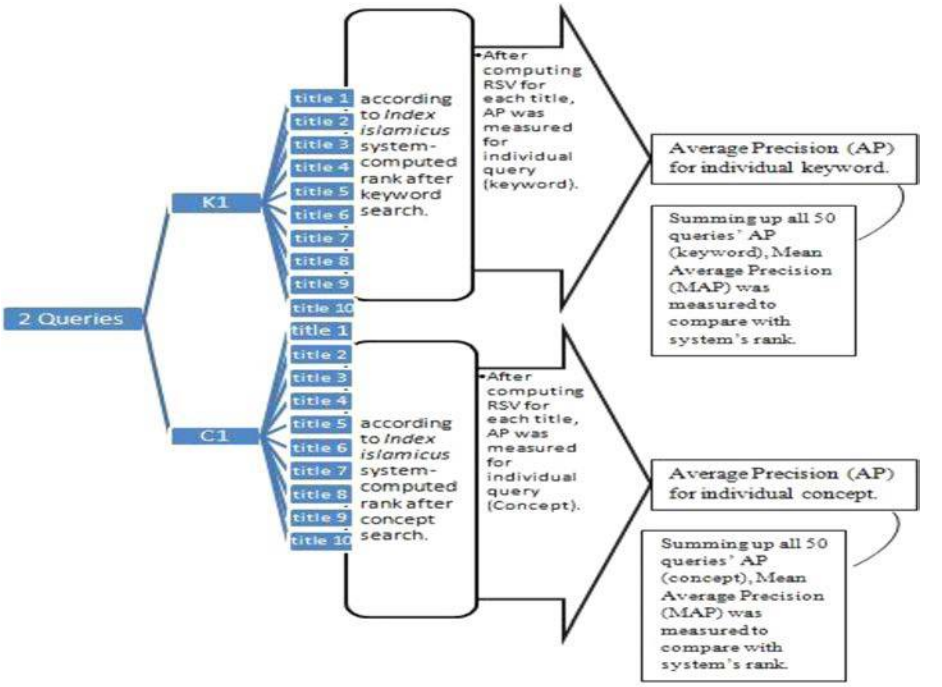


Figure 3.
Example of a query evaluation through keyword and concept

Where *i* is an item from ranked result (first ten) and k is total ranked result (those first ten or less than ten. according to relevance judgment by RSV).

$$\mathrm{AP} = \frac{\text{total RSV of an item (1st + 2nd + 3rd + … 10th)}}{\text{Total items in ranking}}$$

Also, to measure MAP over the set of queries (either keyword or concept):

$$\mathbf{MAP} = \frac{\mathrm{RSVk/c} + \ldots \mathrm{RSVk/c}}{50} \tag{4}$$

k is the keyword and c is the concept (query). 50 is the total queries which is for keyword 50 queries and for concept 50 queries (Table I).

5.3 Analysis of methods

The core 50 titles for query formation were used due to the vast collection of the *Islamic History* in *Index Islamicus*. To generate the notion "‖k‖" (total number of word in a query), Microsoft *Word Count* was used. For stemming purposes, among many other rules, this research combined Paice and Husk's rule with Loving's method of stemming. According to Meadow *et al.* (2007, p. 81), "the more rules of the types are illustrated, the greater the probability of correctly stemming a word". In their rules, the common procedure is to replace last three letters "IES" with Y such as Plaies > Play; omitting last

Symbols	Definitions
d‖∩‖q	Total number of shared elements (keyword/concept word) between a document and a query
N	The entire database size
Ne	Total retrieved result
fd,k	Number of occurrences of a keyword (query) in a document (title + index term)
fd,c	Number of occurrences of a concept (query) in a document (title + index term)
‖d‖	Total number of word in a document (title)
‖k‖	Total number of word in a query (keyword/concept)
MAP	MAP to compare with system computed rank

Note: As keyword search is searched as phrase, the query is considered as one word except ‖k‖. In this case, for example, "legal tradition" is combination of two words, but since its phrase search, in "fd,k" [number of occurrences of a keyword (query) in a document including title and index term] it is considered one word. But in "‖k‖" (total number of words in a query) it breaks into word by word. So "legal tradition in" "fd,k" is one word but in "‖k‖" is two words (Table I)

Table I. Each abbreviated notion indicates to

three letters "ING" such as Singing > Sing and keep at least two letter such as Owing > Ow from last part of the word; omit last part of word when contains "TION" such as Pollution>Pollute and drop the letter "S" when ends with "ES" such as Goes > Go, Places > Place, etc.

After each search, the retrieved result was sorted according to *Index Islamicus relevance* ranking by default.

Example of RSV on *Index Islamicus* according to formula (1) *keyword*:

$$\mathrm{RSVc}\,(\mathrm{d},\mathrm{q}) = \|\mathrm{d}\cap\mathrm{q}\|\,\mathrm{k} \times \left(\sum \frac{\mathrm{N}}{\mathrm{Nk}} \times \frac{\mathrm{fd},\mathrm{k}}{\|\mathrm{d}\|} \times \|\mathrm{k}\|\right)$$

5.3.1 Step 1. Using a random query "q", *Index Islamicus* is searched. The total collection of *Index Islamicus* is about 462,012 which are notated by "N"-the corpus (Figure 4).

The more elements (words/phrases) match with documents, the more relevant result can be delivered.

5.3.2 Step 2. After searching, there are, for example, 40 titles have been retrieved by *Index Islamicus* and they are sorted according to relevance to the query by system's default and ordered in rank. The entire retrieved result is notated with "Nk". Each title is considered as one document and notated with "d". After finding "Nk", the size of entire corpus (collection of Index Islamicus) which is "N", is divided by retrieved result "Nk" (N / Nk) (Figure 5).

$$\mathrm{RSV} = \|\mathrm{d}\cap\mathrm{q}\| \times \left(\sum \frac{\mathrm{N}}{\mathrm{Nk}} \times\right)$$

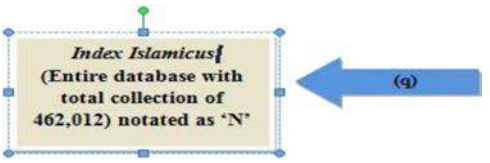


Figure 4. Behind the search (Step 1)

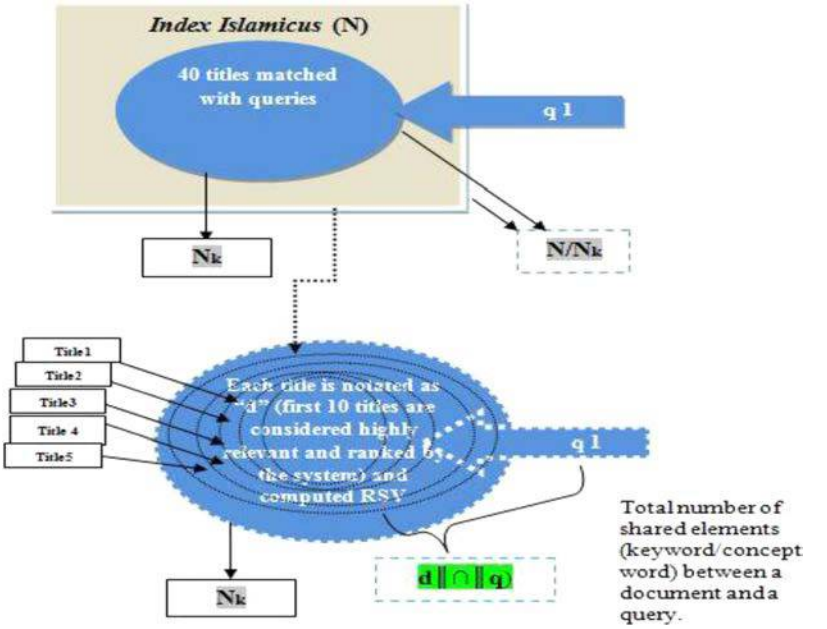


Figure 5.
After search (Step 2)

5.3.3 Step 3. The first title is picked up to calculate RSV, for example. According to the intersection "∩", the common (similar) word with query in title is calculated including query itself as the base of RSV said, "Total number of shared elements between a document and a query", which is notated by "d‖∩‖q" (Figure 5).

5.3.4 Step 4. "fd,k" indicated the occurrences of keyword (the query) in the title as well as inside the index terms/subject heading of Index Islamicus; and"‖d‖" indicated the total number of word in a document which here means title. Finding both "fd,k" and "‖d‖", "fd,k" is divided by "‖d‖" (fd,k / ‖d‖) (Figure 6).

These steps are similar in concept search computation but only in keyword search that number of total words of keyword (query) is also counted such as "‖k‖" indicated the total number of word in a query (keyword).

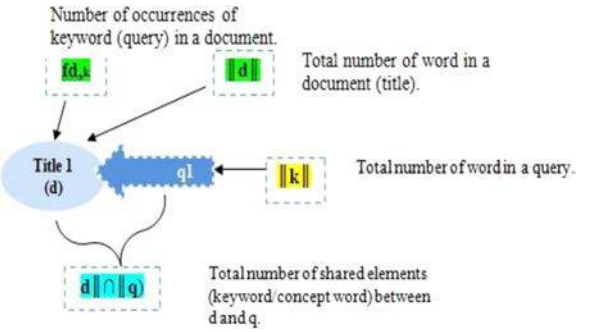


Figure 6.
Output of search (Steps 3, 4 and 5)

5.3.5 Step 5. In the last step, dividing "N" with "Nk" ("Nc" in concept search) and "fd,k" with "‖d‖", the results are multiplied in between such as N / Nk by fd,k / ‖d‖. Then the entire formula from "shared elements" (d‖∩‖q) until "total keyword" (‖k‖) are generated to multiplication, which is d‖∩‖q by ($\sum$N / Nk by fd,k / ‖d‖ by ‖k‖) with percentage such as in Figure 6.

$$\mathrm{RSVc}\,(d, q) = \|d \bigcap q\|\, k \times \left(\sum \frac{N}{Nk} \times \frac{fd, k}{\|d\|} \times \|k\| \right)$$

Then, the result of RSV of both "keyword" search and "concept" search is computed for AP. After AP, combining all queries' AP, the overall MAP is evaluated (Kak, 2013; Cormack and Lynam, 2006).

Basically, the AP is computed through calculating the value of precision and leaving the result that is not relevant. Because this research calculates the RSV, which is another method to calculate recall and precision measuring relevance judgment and accuracy of ranking, RSV was used to find AP, as well as MAP for this research. Examples of AP and MAP calculations are as below.

Evaluating RSV of first ten results after each search and leaving the rest of the results below ten that are considered not eligible to be ranked as most relevant.

The same goes to other 50 queries (concept) (Figure 7).

5.4 Validity and reliability

Abdulahhad and his researchers have proved the validity and reliability measurement of RSV. They tested the model on *ad hoc* image-based retrieval and case-based retrieval where the database contained 230,000 images with English caption and case-based retrieval contained more than 55,000 articles in English. They used 30 queries for image-based retrieval and 10 queries for articles using three different types of indexing

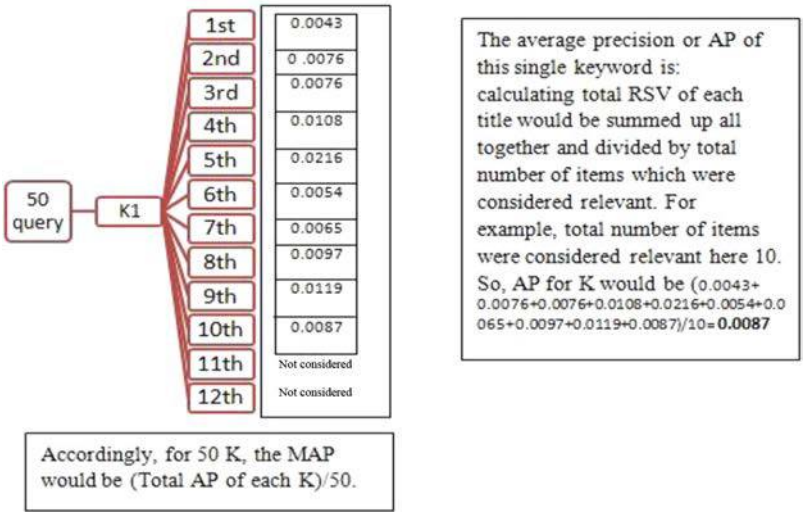


Figure 7.
Example of mean average precision (MAP)

elements: keyword, ngram and concept. Through this performance, their goals were to justify this RSV model itself and study the performance of multiple indexing elements at the same time. The result showed the successful validity and reliability of the RSV model (2011 and 2012).

5.5 Data collection

The actual data for this research were divided into three categories. The first category is the 50 titles of the almost 13,000, which were collected from *Islamic History and Civilizations* of *Index Islamicus*. The second category is formation of those titles into query, based on two indexing elements, which are, "keyword" and "concept"; and the third category is, searching those keywords and queries on *Index Islamicus* and collecting the retrieved results/titles to generate RSV.

The second category data collection involved the use of LCSH through International Islamic University Malaysia Library "classificationweb" and ProQuest thesaurus (subjects) (one of four ProQuest's online thesaurus beside ERIC, LISA and MeSH) to form up proper queries. Based on the ontology-building framework of Salim *et al.* (2010) for Islamic portal, and analyzing both LCSH and ProQuest thesaurus (subjects), nearly matched term and rather broader terms were followed. The flow of query formation process, as shown in (Figure 8):

- For example, the following process is undertaken to form up a query from the title *Islamic studies and the history of religions: an evaluation.* **Title**: *Islamic studies and the history of religions: an evaluation*.
- Predicted word: *Religious study*.

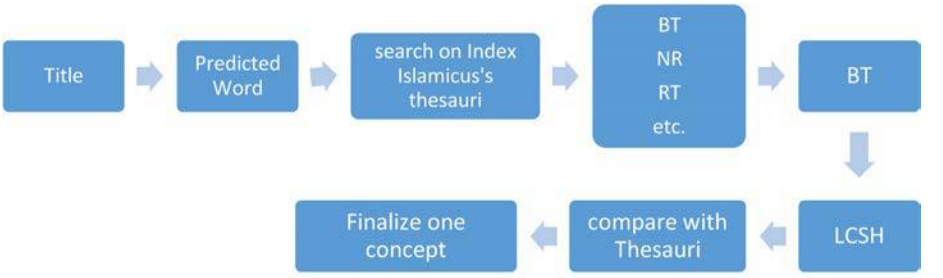
The random word is predicted as main theme of the title.

- Search on *Index Islamicus's* thesauri.

Index Islamicus's thesauri is searched with this predicted word "religious study" to find any comparative word or any other terms relating to this word.

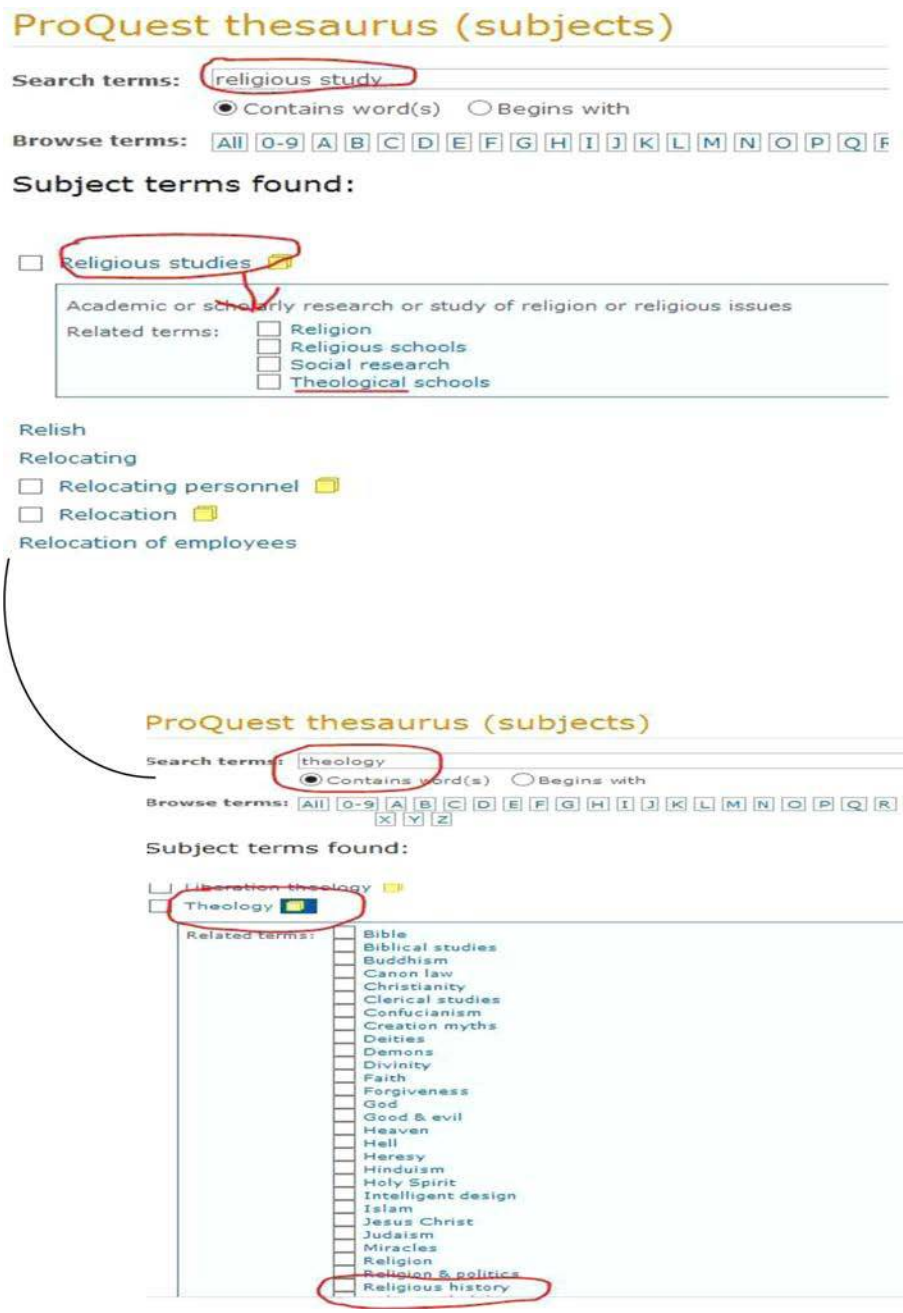
The result shows four different terms under "religious studies".

Skipping other three terms, *Theological schools* are considered as the much related term with *religious studies* because of the common meaning of *theology* as well as to look



Notes: BT = broader term; NR = narrower term; RT = related term

Figure 8.
Example of query formation through LCSH and Index Islamicus thesaurus



for terms related to history because the title's another part is that (*the history of religions*).

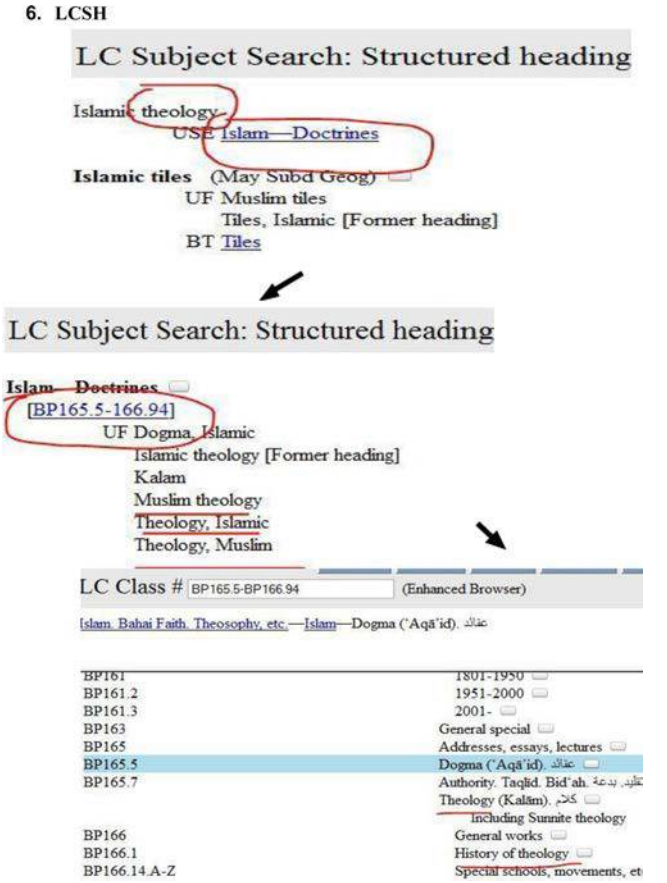
- Religious study > religion, theological schools > (BT, NT, NR).
- Theology > religious history, Islam, religion (BT).

After having broader term *theology*, the thesauri is searched again and finds all other related terms under one umbrella such as *Islam, religion*, *religious history*, etc. under *theology*.

Then, in LCSH and LC class number, *Islamic theology* is found for *Islam-doctrines*. Under *Islam-doctrines*, it is stated that "theology, Islam" and *Muslim theology, History of theology* are the proper terms to refer to *Islamic theology* or *Islamic history*.

- LCSH.

6. LCSH



- History of theology > Islamic theology > Islam > religion > theology.

Hence, comparing all the terms with LCSH, LCC and Index Islamicus's thesauri, the main concept for the title is finalized, which is *theology*.

Concept = *Theology*.
Keyword = *Islamic Study*.

The third category is the research itself. All the retrieved results were snapped shot and the default relevance ranking of IRS (*Index Islamicus*) were recorded for later comparison with RSV. The total titles were 1,000. The RSV for each query, about 10 titles, were computed and 1 AP was calculated. This means, for each query, there would be one AP. So, for the entire queries, 2 MAP would be calculated (50 queries of keyword and 50 of concept).

5.6 Data analysis

The data was analyzed through Microsoft Office Excel. The entire RSV formula was classified for easy calculation. For example, each formula contained about four classes – "total shared elements between *d* and *q*", "total calculation after dividing N by Nk/c", "total calculation after dividing fdk/c by ||d||" and "total words of a document (title) which is ||k||". Thus, the whole formula was calculated using Microsoft Excel (Tables II and III).

6. Findings and analysis

6.1 Keyword search and RSV

After the system's default ranking of each keyword (50) and each concept word (50) search was obtained, the first 10 results were measured according to system's relevance judgment and were given a rank by computing RSV of each title of those first 10. The difference was cleared based on which title among these ten got the highest percentage in relevance criteria by RSV. Then each query's AP was measured. Hence, MAP was generated to compare the relevance status of queries with the documents and system default rank of *Index Islamicus*.

Table IV shows the example of a keyword search and system-computed ranking with RSV evaluation of each ranked title.

The first column indicates the retrieval rank and the last column shows the RSV percentage. The results were sorted according to the highest RSV to the lowest and that made the ranking serial (system's default rank) unordered. So, in Table IV, what was first among the retrieval results in system-computed rank is the sixth according to highest RSV. Thus, fifth, which gained the highest RSV among system-computed rank, has a RSV which is higher than the system's default relevance value or rank. Table V shows all 50 keywords and their RSV with AP.

In Table V, the second column refers to RSV of the first item which was in the first place in ranking order by system's relevance judgment among the first ten items. The

C	D	E	F	G	H	I	J	K	L
d‖∩‖q	N	Ne	**N/Ne = (D/E)**	fd,k	‖d‖	**fd,k/‖d‖ =(G/H)**	**‖k‖**	**C*F*I*J**	**100%/K**

Note: Significance of bold data highlight the formula

Table II. RSV formula in excel for computation (*keyword)*

C	D	E	F	G	H	I	K	L
d‖∩‖q	N	Nc	**N/Nc = (D/E)**	fd,c	‖d‖	**fd,c/‖d‖ =(G/H)**	**C*F*I**	**100%/K**

Note: Significance of bold data highlight the formula

Table III. RSV formula in Excel for computation (concept)

Table IV.
Examples of a keyword search and system-computed ranking with RSV

Keyword ("state-building") Title number in retrieving order	B d‖∩‖q	C *N*	D Nk	E N/Ne = (C/D)	F fd,k	G ‖d‖	H fd,k/‖d‖ = (F/G)	I ‖k‖	J B*E*H*I	K 100%/J (RSV)
Fifth	3	462,012	176	2,625.07	2	16	0.13	1	984.40	**0.1016**
Second	3	462,012	176	2,625.07	2	15	0.13	1	1,050.03	0.0952
Eight	3	462,012	176	2,625.07	2	14	0.14	1	1,125.03	0.0889
Ninth	3	462,012	176	2,625.07	2	10	0.20	1	1,575.04	0.0635
Sixth	3	462,012	176	2,625.07	2	9	0.22	1	1,750.05	0.0571
First	3	462,012	176	2,625.07	2	8	0.25	1	1,968.80	**0.0508**
Fourth	3	462,012	176	2,625.07	2	8	0.25	1	1,968.80	0.0508
Seventh	3	462,012	176	2,625.07	2	8	0.25	1	1,968.80	0.0508
Tenth	3	462,012	176	2,625.07	2	6	0.33	1	2,625.07	0.0381

Query (keyword)	First place in system-computed rank among 10	Highest RSV among ten	AP of each query of the ten relevant result
Keyword 1	0.009	0.641	0.141
Keyword 2	0.037	0.445	0.282
Keyword 3	0.016	0.019	0.012
Keyword 4	0.012	0.028	0.012
Keyword 5	0.018	0.132	0.051
Keyword 6	0.050	0.218	0.146
Keyword 7	0.010	0.084	0.049
Keyword 8	0.004	0.013	0.006
Keyword 9	0.335	2.513	1.395
Keyword 10	0.245	0.245	0.110
Keyword 11	0.004	0.021	0.009
Keyword 12	1.496	5.342	3.160
Keyword 13	0.003	0.008	0.004
Keyword 14	0.206	1.979	0.757
Keyword 15	0.692	2.230	0.784
Keyword 16	0.714	1.489	0.741
Keyword 17	0.020	0.035	0.027
Keyword 18	0.114	0.152	0.103
Keyword 19	0.005	0.007	0.003
Keyword 20	0.107	0.640	0.253
Keyword 21	0.356	1.068	0.615
Keyword 22	0.196	1.223	0.468
Keyword 23	0.051	0.102	0.063
Keyword 24	0.328	1.477	0.827
Keyword 25	0.028	0.178	0.085
Keyword 26	0.034	0.034	0.085
Keyword 27	0.244	0.681	0.416
Keyword 28	0.002	0.026	0.011
Keyword 29	0.531	4.780	2.091
Keyword 30	0.030	0.090	0.054
Keyword 31	0.055	0.055	0.025
Keyword 32	0.096	0.096	0.056
Keyword 33	0.192	0.805	0.403
Keyword 34	0.002	0.006	0.005
Keyword 35	0.014	0.206	0.097
Keyword 36	0.545	4.902	1.786
Keyword 37	0.260	1.671	0.937
Keyword 38	0.149	1.671	0.482
Keyword 39	0.005	0.052	0.020
Keyword 40	0.103	0.239	0.093
Keyword 41	0.028	1.105	0.598
Keyword 42	2.676	13.137	3.976
Keyword 43	0.023	0.121	0.050
Keyword 44	0.004	0.022	0.010
Keyword 45	0.544	0.544	0.341
Keyword 46	0.117	0.571	0.307
Keyword 47	0.234	1.040	0.295
Keyword 48	0.110	0.257	0.075
Keyword 49	0.535	0.257	0.558
Keyword 50	0.056	0.155	0.091

Table V.
The average precision (AP) of each query (keyword) with RSV

third column indicates the highest RSV according to relevance status computation. Then, AP was calculated for each item's average closeness with queries.

6.2 Concept search and RSV

Table IV shows the examples of a concept search and system-computed ranking with RSV evaluation of each ranked title.

Similar to keyword search, in Table VI of concept search, what was the first among the retrieval results in accordance to system-computed rank, is the fifth and thus eighth became the first, achieving highest RSV among system-computed rank. Table VII shows all 50 concept search with MAP computing first 10 results or system's default ranking on the first 10 items.

The experimental results of applying RSV of Abdulahhad *et al.* (2012a, 2012b) to *Index Islamicus* using two types of index elements (keyword and concept) are shown in Table VIII.

Table VIII shows the overall performance of RSV based on two indexing elements: *keyword* and *concept* search (as query). The result of MAP shows that titles which were in the first ten should have the minimum precision rank, that is, 0.45 for keyword and 0.91 for concept search on average. Meaning that, if any item obtained below 0.45 and 0.91 after a query search (keyword and concept respectively), it would have been considered not effective in terms of precision and ranking. Therefore, comparing to system-computed rank, the method used in this study (RSV) found that 22 items of 50-keyword search obtained more than 0.45, whereas the system-computed rank obtained only 9 items over minimum AP (0.45). This makes the method's (RSV) ranking mechanism 44 per cent precise and the system default's ranking 18 per cent in terms of accuracy in retrieving items. This is similar to concept search where the system's default ranking obtained 9 items of 50 over MAP, which is 18 per cent and for RSV, 22 of 50, which is 44 per cent.

7. Discussion and conclusion

7.1 Introduction

Index Islamicus is an online database which focuses on Islamic works and publications and was initiated by the Middle Eastern bibliographers. From the popularity and size of the collections, this database gained so much attention from all over the world. Being an online database, it is crucial that its retrieval or relevance performance is tested (Meadow *et al.*, 2007; Abdulahhad *et al.*, 2012a, 2012b) because an IRS ranks its retrieval results through measuring the closeness of a user's query with its existing documents. The closer a relationship a query gets to a document the higher the rank of that document in retrieved results (Kettani and Newby, 2010) and that makes sense of relevance judgment of an IRS. If this mechanism is not found in any IRS such as in *Index Islamicus*, the RSV of that would be assumed low and weak. The problem is, the performance of *Index Islamicus* on ranking mechanism was not examined by previous researchers directly where only its options, modules and index term descriptors (Kopycki, 2003) were evaluated. Therefore, this study aimed to examine the ranking mechanism of *Index Islamicus* through measuring its one specific subject domain. Hence, beside all other evaluation measure, RSV was selected to test *Index Islamicus* system-computed ranking and compare it with MAP.

Table VI. Example of a concept search and system-computed ranking with RSV evaluation

Concept (theology) Title number in retrieving order	B d‖∩‖q	C *N*	D Nk	E N/Ne =(C/D)	F fd,k	G ‖d‖	H fd,k/‖d‖ = (F/G)	I B*E*H	J 100%/I
Eight	5	462,012	4,273	108.12	4	18	0.22	120.137	**0.8324**
Fourth	6	462,012	4,273	108.12	5	17	0.29	190.806	0.5241
Sixth	5	462,012	4,273	108.12	4	11	0.36	196.588	0.5087
Seventh	5	462,012	4,273	108.12	4	10	0.40	216.247	0.4624
First	7	462,012	4,273	108.12	6	17	0.35	267.129	**0.3744**
Second	7	462,012	4,273	108.12	6	13	0.46	349.322	0.2863
Third	5	462,012	4,273	108.12	4	6	0.67	360.412	0.2775
Fifth	6	462,012	4,273	108.12	5	8	0.63	405.463	0.2466
Ninth	5	462,012	4,273	108.12	4	5	0.80	432.494	0.2312
Tenth	5	462,012	4,273	108.12	4	4	1.00	540.618	0.1850

Table VII.
The average precision (AP) of each query (concept) with RSV

Query (Concept)	First place in system-computed rank among ten	Highest RSV among ten	AP of each query of ten relevant result
Concept1	0.374	0.832	0.393
Concept 2	0.816	1.122	0.524
Concept 3	1.456	1.465	0.827
Concept 4	0.177	0.355	0.155
Concept 5	0.019	0.035	0.017
Concept 6	0.924	3.080	1.584
Concept 7	0.223	0.476	0.233
Concept 8	0.006	0.057	0.025
Concept 9	0.010	0.032	0.017
Concept 10	0.009	0.059	0.036
Concept 11	0.027	0.100	0.037
Concept 12	0.018	0.058	0.027
Concept 13	3.557	18.123	7.055
Concept 14	6.447	9.671	3.744
Concept 15	0.017	0.052	0.031
Concept 16	0.006	0.040	0.018
Concept 17	0.210	0.564	0.286
Concept 18	0.374	3.461	1.487
Concept 19	0.608	0.760	0.296
Concept 20	0.852	3.090	1.082
Concept 21	0.947	0.947	0.587
Concept 22	0.041	0.054	0.030
Concept 23	0.810	5.828	2.009
Concept 24	0.322	1.040	0.427
Concept 25	0.048	0.080	0.046
Concept 26	0.009	0.022	0.012
Concept 27	1.133	3.075	1.089
Concept 28	0.774	0.774	0.404
Concept 29	0.212	2.224	0.998
Concept 30	0.004	0.048	0.020
Concept 31	0.076	0.381	0.211
Concept 32	0.612	2.039	0.777
Concept 33	0.475	0.697	0.478
Concept 34	0.359	0.616	0.392
Concept 35	0.059	0.144	0.071
Concept 36	0.006	0.006	0.004
Concept 37	0.009	0.084	0.031
Concept 38	0.051	0.437	0.156
Concept 39	0.647	1.419	0.587
Concept 40	0.878	1.524	0.867
Concept 41	2.990	19.934	7.635
Concept 42	0.001	0.003	0.002
Concept 43	0.058	0.195	0.064
Concept 44	0.879	2.380	0.892
Concept 45	0.464	1.461	0.699
Concept 46	0.270	0.676	0.279
Concept 47	0.122	0.367	0.279
Concept 48	0.211	0.678	0.341
Concept 49	0.002	0.022	0.008
Concept 50	10.356	14.424	8.465

7.2 Discussion

Through the hypothesis and research objectives, the target of this research was to evaluate *Index Islamicus* subject domain – *Islamic History and Civilizations*-based on how close the correlation of its documents is with queries that are derived from the same domain. According to the assumptions, the more a document and query shared their elements, the more semantically closer they are and the higher they are in retrieval order and obtain better ranking (Abdulahhad *et al.*, 2011; Kettani and Newby, 2010).

Evaluating this assumption through RSV and MAP, this research presented results in previous section and the following is the discussion of the results according to the specific objectives of this research.

Research objective 1: "To determine RSV for Index Islamicus on the subject of Islamic History and Civilizations".

The questions addressed to achieve this aim are as follows:

7.2.1 What is the relevance judgment criteria commonly applied to the subject of Islamic History and Civilizations in Index Islamicus? The relevance judgment in this research referred to the measurement of relevance status of an item. Among many other measurement evaluation methods, RSV method was used to justify relevance judgment criteria of *Index Islamicus* on its subject – *Islamic History and Civilizations*. The judgment criteria obtained after evaluating through RSV is quite common like other online databases, except it has one difference, which is that other databases usually provide full text or abstract, whereas *Index Islamicus* provides only metadata.

In addition, being an IR, *Index Islamicus* relates its documents to the query by estimating the contribution of a query (Abdulahhad *et al.*, 2011) or indexing element such as "keyword" and "concept" in overall matching with the document. As it is a database which provides only citations, this match-making happens with a particular document via its title, index terms and publication titles (Ahmad, 1981; Anwar, 2001). Therefore, the judgment criteria obtained after evaluating this match-making and relevant ranking shows that, ranking mechanism or relevance judgment of *Index Islamicus* on its subject domain – *Islamic History and Civilizations* – are comparatively 18 per cent accurate; which means, accuracy to retrieve exact documents from *Islamic History and Civilizations*, using certain queries of same subject is 18 per cent only.

7.2.2 What are the available controlled vocabulary tools or thesaurus for Islamic History and Civilizations *of* Index Islamicus*?* Like Online Public Access Catalog (OPAC), online database also needs indication linkage or subject access to lead the user or their query into the correct and precise area of interest. OPAC does it through LCSH, where online database does it through building up ontology on specific subject area. Ontology or thesauri for an online database is important because match-making and correlation between query and subject term is decided by it, considering which document should be retrieved and ranked (Huang and Zhang, 2010; Xiaohui *et al.*, 2010).

Type of query	MAP	Highest by RSV	Highest by system	Over MAP by RSV	Over MAP by system rank
Keyword	0.4593	13.137 (K42)	2.676 (K42)	22 out of 50	9 out of 50
Concept	0.9147	19.934 (C41)	10.356 (C50)	20 out of 50	9 out of 50

Table VIII. The total experimental results

Several researches were done on *Index Islamicus*, but none specifically mentioned about its thesaurus. As this is a part of ProQuest, it is assumed that ProQuest's four thesaurus are the overall source of *Index Islamicus*'s thesauri guideline. The four thesaurus are ERIC, MeSH, LISA and ProQuest (subject). As it is predicted, none of this thesauri is particularly specialized for Islamic domain unlike MeSH, which specializes in medical area, and ERIC, which is more focused on education rather religion or Islam in particular (CSA Illumina, 2013).

Research objective 2: "To examine the semantic relationship between documents and queries for the subject of Islamic History and Civilizations in Index Islamicus".

To discuss the second objective, the questions were put as.

7.2.3 What is the RSV for Islamic History and Civilizations *in* Index Islamicus*?*

The RSV is the value of an item to measure its relevance with its query. The *de facto* of an IRS is to retrieve an item and put that in rank measuring the space between the query and the item. If the item gains more relation with its query, the system can compute the relation with a mathematical evaluation and comes up with a result, which is reflected in its retrieval rank. The higher the item appears in the retrieved list, seems the closer its relation with the query.

Hence, the RSV of *Islamic History and Civilizations* of *Index Islamicus* was measured item by item and that was only the first ten items of each search considering those first ten are the highest among all retrieved items. Searching through 100 queries (50 keywords and 50 concepts), the evaluation found that the relevance status is 18 per cent of that subject domain in Index Islamicus. Meaning that, by default, or by system-computed relevance judgment between a query and documents is 18 per cent accurate. In other words, according to hypothesis (the more shared elements a document and a query have, the closer and semantically related they are), 18 of 100 of its retrieved item can match or have perfect correlation with its query. Because this RSV was computed on system's best ten ranked items, it can be said that *Index Islamicus* ranking mechanism is 18 per cent effective; the rest is not that good enough because semantic correlation of documents were not found with query which was proven by RSV evaluation that gained 40 and 44 per cent accuracy.

7.2.4 What are the differences between RSV ranking and Index Islamicus *system's default ranking?*

Mathematically, the difference between *Index Islamicus* system-computed ranking with RSV is 18 of 100 and 44 of 100. The ranking mechanism of *Index Islamicus* works with match-making of query on its metadata unlike other full text or abstract-providing database. In other databases, when a search term (query) happens, the system tries to interact with its index terms as well as documents inside to find the correlation with that search term (query) (Ismail *et al.*, 2000; Kettani and Newby, 2010).

According to RSV measurement in this research, it is found that the more words an item or a document contains, the higher its relevance value will be. For example, Table IX shows fifth gained the highest RSV, where its total words (according to title only because *Index Islamicus* does not provide full text or abstract) is 32, and first, which was first in system-computed rank gained RSV 0.87, contained 15 words in its title, which is a big difference between fifth and first.

Table IX. Differences between RSV ranking and Index Islamicus system's default ranking

Concept (Islamic law) Title number in retrieving order	B d‖∩‖q	C *N*	D Nk	E N/Ne =(C/D)	F fd,k	G ‖d‖	H fd,k/‖d‖ =(F/G)	I B*E*H	J 100%/I
Fifth	9	462,012	2,1650	21.34	7	**32**	0.22	42.013	**2.380203**
Sixth	10	462,012	2,1650	21.34	8	27	0.30	63.230	1.581534
Second	10	462,012	2,1650	21.34	8	16	0.50	106.700	0.937205
First	10	462,012	2,1650	21.34	8	**15**	0.53	113.814	**0.878630**
Eight	8	462,012	2,1650	21.34	6	9	0.67	113.814	0.878630

Similarly, Table X shows because of having bigger title or containing more words, number 1 according to system-computed ranking remained first gaining the highest RSV among all other items' RSV.

However, through RSV formula, the *Index Islamicus* system ranking does not maintain ranking criteria effectively based on the most shared elements between queries and documents which have been proven for obtaining 18 per cent on its ranking mechanism. This is a big difference that the research has found between *Index Islamicus* system ranking and RSV ranking.

7.3 Limitations

Although the study has proven the weakness *of Index Islamicus* ranking mechanism through mathematical evaluation, which is hardly deniable, the study is limited to 100 sampling titles. For each of 100 queries (keyword and concept), only the first 10 items from the system-computed rank (100 × 10) was evaluated for RSV. Therefore, the ranking criterion of *Index Islamicus* was computed only based on its first 10. If number 10 of *Index Islamicus* system-computed rank has gained the highest RSV according to the mathematical evaluation and achieved the first place, then it can be assumed that number 11th or 18th could have the probability of gaining ranking place in the first ten having better RSV. This could be done in future research.

Another limitation of this study was the specific subject area. *Index Islamicus* is a huge database, which contains more than 60,000 entries on the context of Islam and this study focused only on the subject of *Islamic History and Civilizations*. In addition, the study focused on one database which is *Index Islamicus*. If the coverage was wider with higher sampling titles and comparison with other databases was done taking real users' judgment in account, the results might have reflected the impact of the intervention more accurately.

7.4 Recommendations

Based on the results of the study, there are several recommendations that can be proposed for future research. First, future study should include real users' judgment, evaluate other subject area of *Index Islamicus*, comparing *Index Islamicus* with other full-text database or abstract-based database and compute more items than the first ten in the ranking. Second, to improve its index terms variations like unambiguous spelling such as "Medina" or "Madina", "Moslim" or "Muslim" or "Moslem", etc., other index term evaluations like *tf-idf* can be added with RSV evaluation. These spelling variations sometimes make the search or let the system get unwanted ranking list. Finally, this study only measured the ranking mechanism of *Index Islamicus* where future study could evaluate its thesaurus and subject term specifications also.

7.5 Conclusion

Several conclusions can be made from this research based on the results. One of the conclusions is that *Index Islamicus's* RSV is 18 per cent effective in terms of ranking the results on the subject of *Islamic History and Civilizations*. This value may affect the relevance ranking mechanism of this database, which may mislead the user by providing less related item in the rank. Because when a system ranks the retrieved items based on "relevance", it is assumed that those titles that come first on the list are more related with the search query. Hence, this research found that the ranking list provided by *Index Islamicus* for the *subject of Islamic History and Civilizations* after each search

Table X. Differences between RSV ranking and Index Islamicus system's default ranking

Keyword ("Islamic origin") Title number in retrieving order	B d‖∩‖q	C *N*	D Nk	E N/Ne =(C/D)	F fd,k	G ‖d‖	H fd,k/‖d‖ =(F/G	I ‖k‖	J B*E*H*I	K 100%/J
First	2	462,012	27	17,111.56	1	**23**	0.04	2	2,975.92	**0.0336**
Second	2	462,012	27	17,111.56	1	15	0.07	2	4,563.08	0.0219
Eight	2	462,012	27	17,111.56	1	11	0.09	2	6,222.38	0.0161
Tenth	2	462,012	27	17,111.56	1	9	0.11	2	7,605.14	0.0131
Sixth	2	462,012	27	17,111.56	1	7	0.14	2	9,778.03	0.0102

term by its system-computed mechanism was only 18 per cent effective. This means the ranking is not perfectly measured through relevance or closeness of query and document match; or another perception that can be drawn is that because of only citations – not providing full text or abstract – the relevance mechanism with queries did not happen effectively.

The final conclusion is the effectiveness of performance measure of RSV comparing with MAP. Through RSV, this research been able to justify *Index Islamicus* relevance ranking and found actual perfection to make a ranking list based on the hypothesis – more shared elements a document and a query have, semantically closer and higher in rank – is the correlation of a query with its document. Consequently, apart from system-computed rank, this method gained 40 and 44 per cent accuracy or more precise rank for concept and keyword (queries). Accordingly, this evaluation can play a great role for other databases to check and maintain their ranking method. It is certain that a proper rank can be provided by a database when the correlation of user's query with database's existing document would be measured correctly. Therefore, implication of this study could be a guideline for further research on ranking mechanism of other search engines. Researchers and vendors of online databases can ensure their users a true platform of search engine with proper ranking list.

References

Abdulahhad, K., Chevallet, J.P. and Berrut, C. (2011), "Multi-facet document representation and retrieval", in Petras, V., Forner, P. and Clough, P. (Eds), CLEF (Notebook Papers/Labs/Workshop), Amsterdam, ISBN: 978-88-904810-1-7.

Abdulahhad, K., Chevallet, J.P. and Berrut, C. (2012a), "Matching fusion with conceptual indexing", Actes du 4ème Atelier Recherché d'Information SEmantique (RISE), Bordeaux, available at: http://hal.archives-ouvertes.fr/hal-00770561 (accessed 22 March 2013).

Abdulahhad, K., Chevallet, J.P. and Berrut, C. (2012b), "The effective relevance link between a document and a query", in Liddle, S.W., Schewe, K.D., Tjoa, A.M. and Zhou, X. (Eds), *Database and Expert Systems Applications*, Springer Berlin Heidelberg, Berlin, Vol. 7446, pp. 206-218.

Ahmad, A. (1981), "Index islamicus [In Arabic]", *Rissalat Al-Maktaba (The Message of the Library)*, Vol. 16 No. 1, available at: http://search.proquest.com/docview/57029622/6313C892254D45C3PQ/1?accountid=44024 (accessed 22 March 2013).

Ansari, M. (2005), "Matching between assigned descriptors and title keywords in medical theses", *Library Review*, Vol. 54 No. 7, pp. 410-414.

Anwar, M.A. (2001), "Muslims and Europe: a demographic study of citations from the index islamicus database", *Malaysian Journal of Library and Information Science*, Vol. 6 No. 1, pp. 93-104.

Blanke, T. and Lalmas, M. (2011), "Specificity aboutness in XML retrieval", *Information Retrieval*, Vol. 14 No. 1, pp. 68-88.

Chatvichienchai, S. and Tanaka, K. (2009), "An effective document search technique by semantic relationship approach", *Fourth International Conference on Computer Sciences and Convergence Information Technology*, *Seoul*, pp. 53-58.

Chawla, S. and Bedi, P. (2008), "Improving information retrieval precision by finding related queries with similar information need using information scent", *First International Conference on Emerging Trends in Engineering and Technology*, ICETET, Nagpur, pp. 486-491.

Chowdhury, G.G. (2010), *Introduction to Modern Information Retrieval*, Facet Publishing, London.

Chu, H. (2010), *Information Representation and Retrieval in the Digital Age*, Information Today, New Jersey, NJ.

Chunchen, L. and Jianqiang, L. (2012), "Semantic-based composite document ranking", *Proceedings of the 2012 IEEE Sixth International Conference on Semantic Computing, IEEE Computer Society, Washington, DC*, pp. 126-129.

Cormack, G.V. and Lynam, T.R. (2006), "Statistical precision of information retrieval evaluation", *The 29th Annual International SIGIR Conference*, Seattle, Washington, DC, pp. 533-540.

CSA Illumina (2013), "Index Islamicus", available at: www.csa.com/factsheets/islamicus-set-c.php (accessed 12 March 2013).

Dinh, D., Tamine, L. and Boubekeur, F. (2013), "Factors affecting the effectiveness of biomedical document indexing and retrieval based on terminologies", *Artificial Intelligence in Medicine*, Vol. 57 No. 2, pp. 155-167.

Feldvari, K. (2005), "Thesauri usage in information retrieval systems: example of LISTA and ERIC database thesaurus", *2nd International Conference INFuture 2009: The Future of Information Sciences: Digital Resources and Knowledge Sharing*, 4-6 November 2009, *Zagreb*, pp. 279-288.

Huang, G. and Zhang, X. (2010), "Text retrieval based on semantic relationship", *International Conference on E-Product E-Service and E-Entertainment*, ICEEE, 7-9 November 2010, *Henan*, pp. 1-4.

Ismail, A.R., Sembok, T.M.T. and Zaman, H.H.B. (2000), "Search engines evaluation using precision and document-overlap measurements at 10-50 cutoff points", *Proceeding of TENCON*, 24-27 September 2000, Kuala Lumpur, Vol. 3, pp. 90-94.

Kak, A. (2013), "Evaluating information retrieval algorithms with significance testing based on randomization and student's paired t-Test", *Tutorial Presentation*, Purdue University, 7 April.

Kazai, G. and Lalmas, M. (2006), "Inex 2005 evaluation measures [lecture notes in computer science]", in Fuhr, N., Lalmas, M., Malik, S. and Kazai, G. (Eds), *Advances in XML Information Retrieval and Evaluation*, Springer Berlin Heidelberg, Vol. 3977, pp. 16-29.

Kettani, H. and Newby, G.B. (2010), "Instability of relevance-ranked results using latent semantic indexing for web search", *Proceedings of the 43rd Hawaii International Conference on System Sciences, 5-8 January 2010*, *Honolulu*, pp. 1-6.

Kim, G. (2006), "Relationship between index term specificity and relevance judgment", *Information Processing and Management*, Vol. 42 No. 5, pp. 1218-1229.

Kopycki, W. (2003), "Index Islamicus", *Domes*, Vol. 12 No. 1, p. 60.

Manning, C.D., Raghavan, P. and Schutze, H. (2009), *Introduction to Information Retrieval*, Cambridge University Press, Cambridge, MA.

Meadow, C.T., Boyce, B.R. and Kraft, D.H. (2007), *Text Information Retrieval Systems*, Academic Press, San Diego, CA.

Nasraoui, O. and Zhuhadar, L. (2010), "Improving recall and precision of a personalized semantic search engine for e-learning", *Fourth International Conference on Digital Society*, 10-16 February 2010, Sint Maarten, pp. 216-221.

Olson, H.A. and Boll, J.J. (2001), *Subject Analysis in Online Catalogs, Libraries Unlimited*, A Division of Greenwood Publishing Group, Englewood, CO.

Othman, R. (2009), *Trends in Information Retrieval System*, IIUM Press, Kuala Lumpur.

Ouchetto, O., Ouchetto, H. and Roudies, O. (2012), "Relevance ranking for services retrieval", *Journal of Computer Science*, Vol. 8 No. 10, pp. 1667-1673.

Park, L. (2011), "Bootstrap confidence intervals for mean average precision", *Proceedings of the Fourth Annual ASEARC Conference*, 17-18 February 2011, University of Western Sydney, available at: http://ro.uow.edu.au/asearc/22/ (accessed 12 March 2013).

Radovanović, M. and Ivanović, M. (2008), "Document representations for classification of short web-page descriptions", *Yugoslav Journal of Operations Research*, Vol. 18 No. 1, pp. 123-138.

Salim, J., Farhana, S., Hashim, M. and Aris, A. (2010), "A framework for building multilingual ontologies for Islamic portal", *International Symposium in Information Technology (ITSim)*, 15-17 June 2010, Kuala Lumpur, pp. 1302-1307.

Sato, H. (2007), "The main purpose and the basic rules of developing ExpressFinder/Thesaurus", *Journal of Information Science and Technology Association (Joho no Kagaku to Gijutsu)*, Vol. 57 No. 2, pp. 84-88.

Serbanoiu, A. and Rebedea, T. (2013), "Relevance-based ranking of video comments on youtube", *19th International Conference on Control Systems and Computer Science, CSCS*, 29-31 May 2013, Bucharest, pp. 225-231.

Xiaohui, T., Yuefeng, L., Lau, R.Y.K. and Geva, S. (2010), "Ontology-based specific and exhaustive user profiles for constraint information fusion for multi-agents", *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 31 August-3 September 2010, Toronto, pp. 264-271.

Yisong, Y., Finley, T., Radlinski, F. and Joachims, T. (2007), "A support vector method for optimizing average precision", *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 23-27 July 2007, *New York, NY*, pp. 271-278.

Further reading

Brill Academic Publishers (2013), "About index islamicus", *ProQuest LLC*, Leiden, available at: http://search.proquest.com/indexislamicus/productfulldescdetail?accountid=44024 (accessed 12 March 2013).

Robertson, S.E., Kanoulas, E. and Yilmaz, E. (2010), "Extending average precision to graded relevance judgments", *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 19-23 July, 2010, *Geneva*, pp. 603-610.

Corresponding author

Roslina Othman can be contacted at: roslina@iium.edu.my