

# DBIScholar: An iPhone Application for Performing Citation Analyses

Andreas Robecke, Ruediger Pryss, and Manfred Reichert

Institute of Databases and Information Systems, Ulm University, Germany  
{andreas.robecke, ruediger.pryss, manfred.reichert}@uni-ulm.de

**Abstract.** DBIScholar is a free iPhone App that allows for the retrieval and analysis of academic citations. As raw input DBIScholar uses data from Google Scholar. Based on their analysis, a number of citation metrics (e.g., h- and g-index, total number of citations) is calculated. Results are available on screen, but can be also stored and used by other Apps (e.g., email). We believe that DBIScholar and its services will be useful for authors to track the evolution of their citation metrics. Future releases will also cover other mobile platforms (e.g., Android).

**Keywords:** Citation Metrics, iPhone App, H-Index, G-Index

## 1 Introduction

Smart phones have become an indispensable work equipment for many people. As technology is evolving, CPUs have become faster, memory larger and batteries more efficient. Techniques like UMTS as well as the increasing coverage of WLAN access points provide fast mobile Internet connections and enable mobile client applications to communicate with servers located anywhere. Being equipped with a GPS unit together with a respective framework further allows for localisation and navigation functionalities within mobile applications. In the meantime smart phones have proven to be a useful platform for everyday applications. However, it still has to be proven whether contemporary smart phone technology is ready for enabling more sophisticated business applications. We evaluated this by realizing advanced applications on the iPhone as one of the most advanced smart phone platforms currently available. An important issue was to explore the capabilities and restrictions existing for iPhone application development. To elaborate on this we developed the DBIScholar iPhone App for calculating citation indices based on Google Scholar data. Since the application turned out to be more mature than a prototype and also provides interesting features for the scientific community we have decided to make it available for free. It can be downloaded from Apple's App Store [4].

The major DBIScholar feature presented in this paper is to calculate two scholarly indices, namely the *h*- and *g*-index. Both indices aim at measuring the productivity and impact of the work published by a scholar. They are based on the number the top most cited publications are referenced by other papers.

The h-index was defined by J. E. Hirsch in his paper "An index to quantify an individual's scientific research output" [11] in 2005 as the maximum number  $n$  of papers with citation numbers  $\geq n$ . The g-index, in turn, was introduced by Leo Egghe in his paper "Theory and practice of the g-index" [9] in 2006 as an improvement to the h-index. The g-index is defined as the unique number such that the  $n$  most cited articles (together) received at least  $n^2$  citations. The definition of the g-index inherits most properties of the formula of the h-index. In addition, it better takes into account the (few) very best cited articles of an author. The h-index is robust in the sense that it is insensitive to lowly cited papers as well as outstanding highly cited papers. Egghe claims the latter to be a drawback as the evolution of the most cited papers is not being taken into account at all. Once a paper is selected to belong to the top  $h$  papers, it does not influence the calculation of the h-index in subsequent years even if it doubles its number of citations [13]. However, the h-index still seems to be the most popular metrics used for citation analysis.

There are online resources listing the h-indices of the best scholars in their field such as the website "The h Index for Computer Science" [14] and "Arnetminer" [1]. To calculate the h-index of a scholar accurately, it is necessary to know the exact citation counts of her publications. There are a few comprehensive data sources such Web of Science [7], Scopus [6] and Google Scholar [2], which can provide this data in a more or less accurate manner. [8] and [10] discuss and compare these sources in the context of calculating h- and g-indices. Calculation results definitely vary depending on the data source used, and different opinions exist which data source serves best for such calculations. Due to the fact that Google Scholar is the only data source freely available, we use it for our application. As an advantage search results can be reproduced by anyone.

The remainder of this paper is organized as follows. Section 2 gives a short overview of the DBIScholar architecture. Section 3 describes its main features in detail. Finally, Section 4 gives an outlook on future work.

## 2 DBIScholar Architecture

As aforementioned our original goal was to analyse smart phone technologies in terms of business capabilities. Therefore, we considered the idea to develop an application for calculating scholarly indices to be an appropriate task as it involves data retrieval, processing, storage, and visualization. To study as many aspects of iPhone application development as possible and to gain experience on the performance and capabilities of contemporary devices, we decided to implement all functionality within the application instead of outsourcing parts to a server. The flow chart from Fig. 1 illustrates the workflow comprising data retrieval, processing and storage in DBIScholar.

To calculate the h- and g-indices of a particular author, the application's initial interface presents a form to enter the scholar of interest. An advanced form additionally allows the user to specify further details such as publication dates and subject areas similar to the advanced search form offered by Google

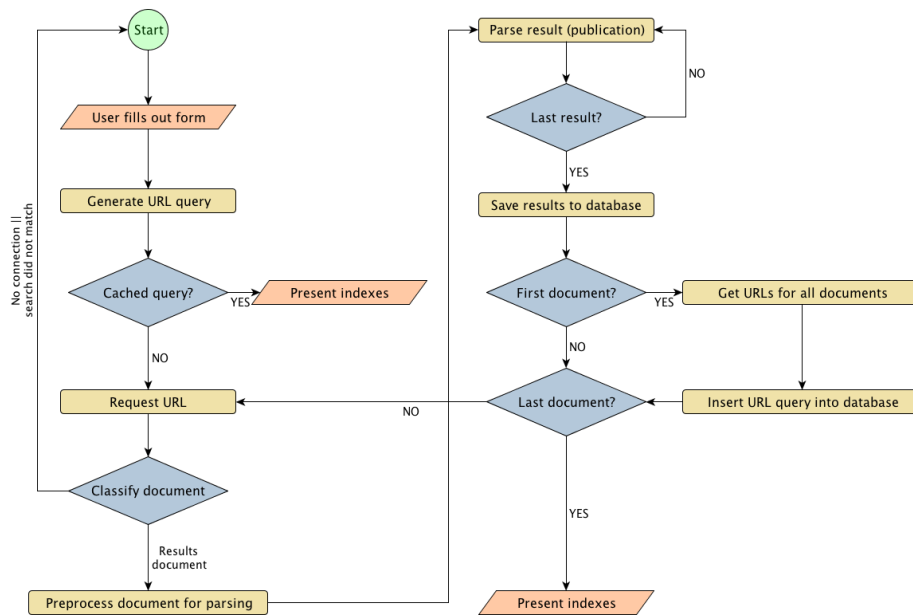


Fig. 1: DBIScholar Search Workflow

Scholar [3] and tools like Publish & Perish [5]. After specifying an author and submitting the filled search form, a respective URL is created to request the corresponding document containing the data of interest from Google Scholar. In the given context this data includes all citation counts concerning the publications of the specified scholar. If there is no cached data, the application requests the document from the created URL. Otherwise, the indices are calculated based on cached data and are immediately presented to the user. If the search returns a valid document it is pre-processed for parsing. In this context we had to consider that any document provided by Google Scholar may contain a maximum of one hundred results (publications). If a scholar has published a higher number of papers, it becomes necessary to request further documents after having parsed the first one. Additionally, to implement the caching mechanism, the query URL has to be saved after parsing the first document.

### 3 DBIScholar Features

**Index Calculation.** DBIScholar features are structured into three tabs. The left tab is the one initially presented to the user after having launched the application. Its root view presents a form allowing the user to search for the publications of a specific scholar (cf. Fig. 2a). If the search matches any result the publication counts of all retrieved publications are determined, the h- and g-indices are calculated, and the obtained results are presented to the user (cf.

Fig. 2b). The user then has the options to save the results, to email them (including all publications the results are based on), or to display the publications.

**Deleting and Merging Publication Entries.** To be able to calculate the indices of a scholar as precise as possible, it is necessary to check whether all publications retrieved during the search actually belong to the scholar of interest. In this context all papers actually contributing to the h- and g-indices are marked in yellow and green colours (cf. Fig. 2c). Thus it is essential to ensure that all these publications belong to the scholar of interest and not to another one with similar name or search criteria.

To allow users to discard publications of "wrong" authors, DBIScholar provides the functionality to delete publications from the list by swiping the finger across it and pressing the appearing "Delete" button. Deleted publications are added at the end of the list to the "Rejects" section. Deleting publications from the "Rejects" section, in turn, allows restoring them. DBIScholar assumes that publications can be uniquely identified by their titles. Occasionally, it happens that Google's search engine delivers multiple publications with the exact same title as result of a search. It may also happen that the same publication is listed multiple times because it was published under different titles or in different form (e.g. as journal paper and a technical report). All cases in which the same publication is listed multiple times are undesirable and are most likely caused by "confusion" of the search engine. In such a case the items should be merged into a single representative publication, and all citation counts be added up. Generally, it is up to the user to figure out which publications are the same. DBIScholar then allows her to merge the identified publications using the "Merge" button in the navigation bar of the *publications list view* (cf. Fig. 2c). In DBIScholar it is possible to merge any publication with others. The publication with the highest citation count is then kept as the representative publication and is marked with "Merge #count" in red colour in the publications list view (where count corresponds to the number of papers merged with this publication).

**Advanced Search Form.** In the initially presented root view (cf. Fig. 2a), the user has the option to extend the form by tapping "Advanced Search...". Amongst others, the advanced search form allows users to specify a subject area or to limit the publications to a certain period of time. This can be especially useful when search results include many publications of an author or multiple authors with similar names. Instead of deleting all the publications of "wrong" authors, it can be an option to use the advanced search to retrieve less non-fitting publications.

**Displaying Publications.** After selecting any publication in the list view, DBIScholar provides further details on it. The appearing user interface offers up to four additional functions. First, if a link to the publication file is available the respective document can be displayed. Second, an "Email Link" button allows

users to send this link via email. Third, if the publication has been merged with others, the "Dissolve" button displays a user interface which allows demerging existing items. Fourth, statistics on citations of the respective paper can be displayed (cf. Sect. 3).

**Reuse Settings of a Previous Search.** Both the search form and the advanced search form allow users to re-apply the settings they have made in the context of a previous query (e.g., to merge certain publications or to delete them). For this purpose DBIScholar allows them to turn on this feature before submitting a query. The App then checks whether the user has submitted the same query before. In this case, it will look for all settings the user has performed on the publications list of her previous search. These are then applied to the publications list obtained as result of the current query as well. However, there are cases for which this feature does not perform well. Occasionally, a single query delivers multiple publications with the exact same title. As another exception a search request might not deliver exactly the same title for an identical publication as another search. Actually, this happens in rare cases. DBIScholar then is unable to figure out that the two publications are actually the same.

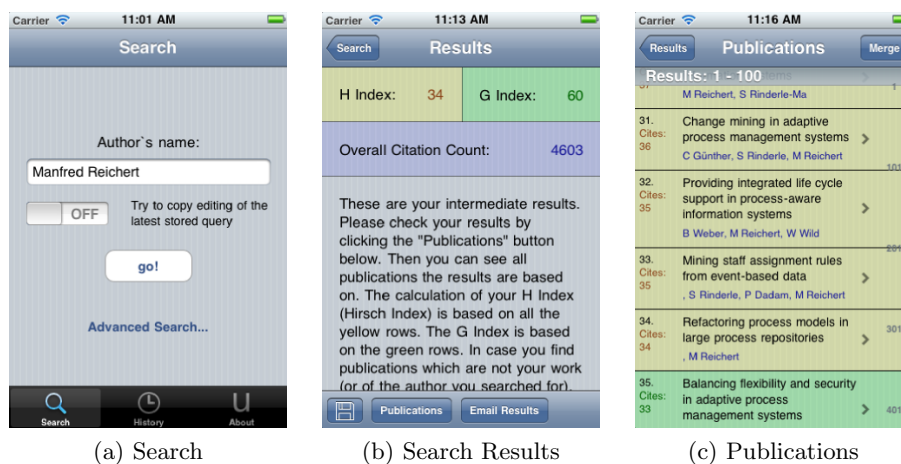


Fig. 2: Search Tab

**Managing Search Results.** The history of all conducted and stored queries may be displayed by tapping the "History" tab on the root view of DBIScholar (cf. Fig. 3a). The history list of previous queries then differentiates between cached and saved queries indicated either by a transparent or a shiny blue floppy disk (cf. Fig. 3a). A history entry itself displays the author's name and the date

of the query. When selecting an entry from the history list, the displayed user interface is exactly the same as the one obtained when initially triggering a search. Thus all features introduced so far are also accessible within the "History" tab. The only difference is that the results view, which displays the calculated indices, provides further details of the queries instead of instructions on how to edit the results after initially triggering a query.

**Comparison.** The root view of the history tab provides another interesting feature offered by the "Compare" button in the navigation bar (cf. Fig. 3a). After pressing this button the user may select two queries from the history to compare their results with each other. Doing so the two indices for both queries are displayed in the same screen (cf. Fig. 3b). The two colours yellow and green, which are also used to differentiate between publications solely contributing to the h-index and those contributing to the g-index, are now being used to differentiate between the two results (cf. Fig. 3b). Again, it is possible to display the publications the results are based on. The publications of both searches are then displayed in the same list (cf. Fig. 3c). The colours indicate to which of the two searches a publication belongs. If one publication belongs to both search results, it is displayed in neutral colour (i.e. blue). To compare the citation counts of the two searches anyway, these are displayed in a small section on the left. The citation count of the publication belonging to the second search (green) is being displayed in relation to the first one and thus may display values such as +1, -3, or 0. Note that this feature is not only interesting to compare two scholars, but also to track changes of the citation counts of publications belonging to the same author and their impact on the indices over time. This can be done by submitting the same search once in a while. When comparing the latest search with a previous one, users can find out whether citation counts of their papers have changed or whether this led to updated indices.

**Graph Feature.** DBIScholar allows users to visualize the evolution of the citation counts of a particular publication over time. The respective graph is based on all publications citing this publication. Particularly, it enables predictions on the future evolution of the citation counts. Users have the option to switch between a detailed view (cf. Fig. 4a) providing the exact citation counts per year, and another progression graph (cf. Fig. 4b) showing the progression of the citation counts on a single screen. Since Google Scholar does not always provide a publication date, accuracy of the graphs varies. A percentage icon on top of the graph view indicates (cf. Fig. 4a) its accuracy in percentage based on the total publication count allocated in the graph. By clicking the percentage icon, this exact issue is described in a separate screen as illustrated by Fig. 4c. Finally, it is also possible to send an image of the graph to any email address.

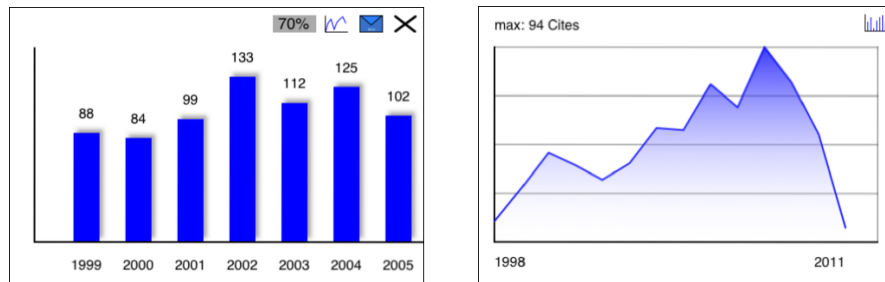


(a) History Root View

(b) Index Comparison

(c) Citation Comparison

Fig. 3: History Tab



(a) Detailed Graph

(b) Progression Graph

This graph visualizes the evolution of the citations over the years. 1365 of a total of 1870 citations could be allocated in this Graph. The accuracy of the graph therefore is about 72% based on the data provided.

(c) Graph Accuracy

Fig. 4: Graph Views

## 4 Outlook

Since it turned out to be a mature application providing useful features for researchers, we will invest further efforts to enhance DBIScholar. One idea is to visualize the evolution of the indices based on the dates of all publications citing an author of interest. This feature would be analogue to the one described in Section 3 for the citation counts of one particular paper. Another extension will be the integration of an offline access feature allowing users to download publications to access them at any time. Some articles about the h index claim that it is possible for authors to manipulate their indices by citing their own publications [12]. Thus, another interesting feature is to count the number of self-citations and to calculate the percentage based on the overall citation count. To further improve DBIScholar and to extend it with additional features we would appreciate getting feedback via Apple's App Store rating system [4] or email. Finally, we will provide DBIScholar on other mobile platforms (e.g. Android and Windows Phone 7) in future.

## References

1. Arnetminer. URL <http://www.arnetminer.org/>.
2. Google scholar. URL <http://scholar.google.com/>.
3. Google scholar - advanced scholar search. URL [http://scholar.google.de/advanced\\_scholar\\_search?hl=en](http://scholar.google.de/advanced_scholar_search?hl=en).
4. itunes dbischolar app store link. URL <http://itunes.apple.com/us/app/dbisscholar/id401979619?mt=8&ls=1>.
5. Publish or perish software. URL <http://www.harzing.com/pop.htm>.
6. Scopus. URL <http://www.scopus.com/>.
7. Web of science. URL <http://isiwebofknowledge.com/>.
8. Judit Bar-Ilan. Which h-index? – a comparison of wos, scopus and google scholar. *Scientometrics*, Vol. 74, No. 2 (2008) 257–271, 2007.
9. Leo Egghe. Theory and practise of the g-index. *Scientometrics*, 2006.
10. Prof. Anne-Wil Harzing. Google scholar - a new data source for citation analysis. 2007.
11. J E Hirsch. An index to quantify an individual's scientific output. *Proc. Natl. Acad. Sci. U. S. A.*, 2005.
12. Konstantin V. Krutovsky Lev A. Zhivotovsky. Self-citation can inflate h -index. *Scientometrics*, 2008.
13. Lokman I. Meho and Kiduk Yang. A new era in citation and bibliometric analyses: Web of science, scopus, and google scholar. *Journal of the American Society for Information Science and Technology*, 2006.
14. Jens Palsberg. The h index for computer science. URL <http://www.cs.ucla.edu/~palsberg/h-number.html>.
15. Andreas Robecke. Development of an iphone business application - diploma thesis, university of ulm. 2011.