Assessing the scholarly impact of evaluation campaigns

Applicant
Dr. Theodora Tsikrika
Post-doctoral researcher
University of Applied Sciences Western Switzerland (HES-SO)
Sierre, Switzerland
theodora.tsikrika@acm.org

Host
Dr. Birger Larsen
Associate Professor
Royal School of Library and Information Science (RSLIS)
Copenhagen, Denmark
blar@iva.dk

1   Purpose of the visit

Evaluation campaigns have been widely credited with contributing tremendously to the advancement of information access systems by (i) providing the infrastructure and resources that support researchers in the development of new approaches, and (ii) promoting the exchange of ideas. Over the years, several large-scale evaluation campaigns have been established at the international level, where major initiatives in the field of textual information retrieval include the Text REtrieval Conference (TREC), the Cross-Language Evaluation Forum (CLEF), the INitiative for the Evaluation of XML retrieval (INEX), and the NTCIR Evaluation of Information Access Technologies. Similar evaluation exercises are also carried out in the field of visual information retrieval, with TREC Video Retrieval Evaluation (TRECVid), PASCAL Visual Object Classes challenge, MediaEval, and ImageCLEF being among the most prominent.

Measuring the impact of such benchmarking activities is crucial for assessing which of their aspects have been successful, and thus obtain guidance for the development of improved evaluation methodologies and information access systems. Given that their contribution to the field is mainly indicated by the research that would otherwise not have been possible, it is reasonable to consider that their success can be measured, to some extent, by the scientific impact of the research they foster. The scientific impact of research is commonly measured by its scholarly impact, i.e., the publications derived from it and the citations they receive.

The purpose of this visit is to work towards methodologies that measure the scholarly impact of evaluation campaigns, not only those based on the traditional system-oriented evaluation paradigm, but also those that will be based on the new evaluation paradigm introduced by ELIAS. To this end, HES-SO has already performed a preliminary assessment of the scholarly impact of ImageCLEF [TGM11], the cross–language image retrieval annual evaluation campaign that was introduced in 2003 as part of CLEF [MCDC10] . RSLIS's expertise in the field of informetrics (including bibliometrics and scientometrics) and in the evaluation of research quality will contribute towards establishing a workflow that will automate and consolidate the methodology employed in [TGM11]. This methodology will then be applied in order (i) to perform an extended scholarly impact analysis for ImageCLEF and (ii) to assess the impact of the whole of CLEF.

2   Description of the work carried out during the visit

The scholarly impact of any *research activity* is commonly measured by the publications associated with it (i.e., the publications generated as a result of this activity) and the citations they receive. Existing work in the area of bibliometrics and scientometrics has focussed on assessing the scholarly impact of specific publication venues (e.g., journals, conference proceedings, etc.) or of the research activities of individual authors, institutions, countries, or entire research domains. Our aim is to establish a workflow and develop

a methodology for assessing the scholarly impact of evaluation campaigns that will also be generic enough to be applicable to any type of research activity associated with a dataset of publications, such as those derived within the context of a particular research project, e.g., ELIAS. This has the potential to have a high impact in providing guidance to different entities for planning their research policy; such entities may lie at various levels of granularity, ranging from single research groups or institutions to nationwide or Europe-wide funding bodies, such as ESF.

During the visit, Dr. Tsikrika and Dr. Larsen discussed extensively and exchanged information on their past work and experience on impact analysis with the goal to first establish a common ground and then to synthesise the workflows each had applied in the past towards a consolidated methodology. In particular, Dr. Tsikrika presented HES-SO's preliminary assessment of the scholarly impact of the ImageCLEF evaluation campaign [TGM11] both in a talk to RSLIS (see Section 6.1) and in private discussions with Dr. Larsen. Dr. Larsen presented to Dr. Tsikrika RSLIS's activities in the field of "*Research Policy*", a focus area concerned with theories, methods and indicators for science studies and research analysis within given subjects. This was followed by a detailed presentation of two reports co-authored by Dr. Larsen on evaluating the impact of Danish research in Natural Sciences [IL07] and in the Information, Communication and Technology sector [SLL08], respectively; these reports had been commissioned to RSLIS by the Danish government.

The discussions between Dr. Tsikrika and Dr. Larsen resulted in the identification of the following three main steps in the process for assessing the scholarly impact of a research activity:

1. Publication data collection
2. Citation data collection
3. Data analysis

Sections 2.1-2.3 describe each of these steps, respectively, present the strategies applied in the past, and outline the decisions that were made and the solutions that were proposed during the visit so as to address the challenges that were encountered.

2.1 Publication data collection

The first step for assessing the scholarly impact of an evaluation campaign is to identify the publications associated with it and collect them in a dataset so that their citation data can then be obtained and analysed. An examination of the publications generated as a result of benchmarking activities indicates that there are typically four main types of such publications (see also [TGM11] for a discussion):

1. *WN:* publications in the Working Notes (Notebooks) accompanying the workshop organised by each evaluation campaign,
2. *Proceedings*: publications in post-workshop Proceedings (if any),
3. *Resources*: publications describing the resources (e.g., test collections, evaluation metrics, etc.) developed in the context of evaluation campaigns; these are typically written by the organisers/coordinators of the campaign and published in venues (e.g., journals, conferences, and workshops) outside the context of the campaign, and
4. *Other*: publications where resources developed in the context of the evaluation campaigns are employed for evaluating the research that is carried out; these are typically published in venues (e.g., journals, conferences, and workshops) outside the context of the campaign.

In ImageCLEF and all other CLEF Labs, publications of all the above types are being generated[1]. In other evaluation campaigns, such as TREC and TRECVID, there are no post-workshop Proceedings, but all other types of publications are encountered.

The complete lists of the *WN* and *Proceedings* publications can be automatically obtained from bibliographic data sources, such as DBLP; the rest need to be discovered. In [TGM11], HES-SO analysed the ImageCLEF *Proceedings* and *Resources* publications. The former were obtained from DBLP by specifying

---

1   To be accurate, this publication scheme was followed until 2009; in 2010 the format of CLEF changed and the there are no longer any follow–up CLEF proceedings, just the Working Notes.

CLEF as the publication venue, while the latter were manually identified given that their authors (ImageCLEF organisers) are known. Our goal is to develop a methodology for obtaining all (to the extent possible) publications associated with an evaluation campaign in an automated (or semi-automated) manner.

Following extensive discussions, Dr. Tsikrika and Dr. Larsen agreed upon a publication data collection methodology consisting of the following steps:

1. Construct an initial "clean" (i.e., manually validated) set of publications D associated with an evaluation campaign.

2. Identify candidate publications to be added to D; the candidate set C is obtained automatically using a bibliographic and citation data source, such as Google Scholar (see Section 2.2 for a discussion):
    - Add to C the publications that are retrieved when querying the data source using the name and/or the URL of the evaluation campaign (e.g., for the case of ImageCLEF use "imageclef", "www.imageclef.org", etc. as queries).
    - Add to C the publications that cite those in D.

3. Eliminate duplicates in C and remove from C those already in D.

4. Validate the publications in C. To determine whether the publications identified in the previous step can indeed be considered for inclusion in D, i.e., that they are actually associated with and derived from the research activities of an evaluation campaign, rather than simply mentioning and/or citing the evaluation campaign in passing, a validation step is required. This validation is typically performed manually by an expert in the field when a relatively small number of candidate publications is involved. Our goal is to investigate methods for automating this process (to the extent possible) by employing machine learning techniques to automatically determine whether a given article contains text that clearly states whether resources of an evaluation campaigns were built or used. To identify these evaluation campaign resources use statements in the candidate publications, an approach similar to that proposed in [NWL11] will be explored.

5. Enlarge D by adding the validated publications.

6. Repeat steps 2-5 until no new publications are added.

This iterative process will be applied for constructing each evaluation campaign-specific dataset. As a first step, the ImageCLEF dataset will be constructed by expanding the one built in [TGM11].

2.2 Citation data collection

The most comprehensive sources for citation data are:

1. commercial providers: Thomson Reuters' Web of Science (generally known as ISI Web of Science or *ISI*), established by Eugene Garfield in the 1960s, and *Scopus*, introduced by Elsevier in 2004, and

2. freely available: *Google Scholar* (GS), launched in 2004, and *Microsoft Academic Search* (MAS), launched in 2009.

Each of these sources follows a different data collection policy that affects both the publications covered and the number of citations found. ISI has a complete coverage of more than 10,000 journals going back to 1900, but its coverage of conference proceedings or other scholarly publications, such as books, is very limited or non-existent. For instance, in the field of computer science, ISI only indexes the conference proceedings of the Springer Lecture Notes in Computer Science and Lecture Notes in Artificial Intelligence series. The citations found are also affected by its collection policy, given that in its General Search, ISI provides only the citations found in ISI-listed publications to ISI-listed publications. Scopus aims to provide a more comprehensive coverage of research literature by indexing nearly 18,000 titles from more than 5,000 publishers, including conference proceedings and "quality web sources". In its General Search, it lists citations in Scopus-listed publications to Scopus-listed publications from 1996 onwards. GS, on the other hand, has a much wider coverage since it includes academic journals and conference proceedings that are

not ISI- or Scopus-listed, and also books, white papers, and technical reports, which are sometimes higly cited items as well. MAS is still a relatively nascent offering to the community, but aims to provide a wide coverage (similar to GS) and its recent efforts in substantially increasing the number of publications it indexes indicates its potential [B11].

As it is evident, these differences in coverage can enormously affect the assessment of scholarly impact metrics; the degree to which this happens varies among disciplines [B08, H10]. In his past work, Dr. Larsen had used ISI for analysing research in Natural Sciences [IL07] and Scopus in ICT [SLL08]. For Computer Science (and ICT in general), where publications in peer-reviewed conference proceedings are highly valued and cited in their own right, without necessarily being followed by a journal publication, ISI greatly underestimates the number of citations found [RT05, B08], given that its coverage of conference proceedings is only very partial, and thus disadvantages the impact of publications. Therefore, it was decided to not consider ISI any further for the assessment of the impact of evaluation campaigns. HES-SO's preliminary assessment of the scholarly impact of ImageCLEF [TGM11] indicated that Scopus has a very limited coverage compared to GS: 249 ImageCLEF publications obtained 2,147 citations in GS and 303 in Scopus. Therefore, it was also decided to not consider Scopus any further. As a result, it was decided to employ the following two sources of citation data: (i) GS due to its wide coverage, and (ii) MAS due to its potential in achieving a wide coverage. MAS was also selected so as to compare and contrast these two data sources; such a comparison has not been previously performed in the context of such an analysis.

Once the citation data sources have been selected, the next step is to query them using the publication data as input so as to obtain the citation data. Whereas ISI, Scopus, and MAS all provide an API, GS does not; this requires the development of effective querying and matching strategies for determining the citations received by a given set of publications. In particular, for each of the publications in a given dataset (defined in the first step; see Section 2.1), a list of relevant GS entries should be retrieved and the one(s)[2] corresponding to the given publication should be identified and their citation data obtained.

In [TGM11], HES-SO collected the citation data by using the Publish or Perish[3] (PoP) system, a software wrapper for GS that supports faceted search over a number of fields[4]. In particular, the following querying strategies were employed for the publications considered:

- *Proceedings* publications: the CLEF proceedings title was used in the Publication field, "image" in the Keywords field, and the publication year of the proceedings in the Year field.

- *Resources* publications: their title and first author name were used in the respective fields. Given the small number of such publications (54) in the preliminary study, a query was manually submitted for each publication.

In both cases, the results were manually refined so as to remove irrelevant entries and merge equivalent ones.

However, the above querying strategy would require extensive manual work and is thus unsustainable for arbitrary datasets that contain publications without common characteristics. For instance, while collecting the citation data for a given publication venue or a given author requires a single query, obtaining the citation data for a set such as the *Resources* or the *Other* publications requires a separate query to be submitted for each publication. Therefore, it was decided to build a wrapper around GS that would automatically query it so as to retrieve a list GS entries matching the publications in any given dataset. For each publication, a query will be submitted to GS, the results will be obtained and parsed so as to identify the relevant GS entries, their citation data will be obtained, and (in case of multiple GS entries corresponding to a single publication) the citation data would be merged. For each step, the following strategies will be applied:

1. *Querying*: For each publication, two types of queries will be submitted: one corresponding to the full title and one consisting of the most informative terms in the title identified using tf.idf. This will result in a list a GS entries that will be fed to the next step.

---

2 GS frequently has several entries for the same publication, e.g., due to misspellings or incorrectly identified years, and therefore may deflate its citation count [RT05, P06]. This though can be rectified through the (manual or automatic) merging of entries deemed to be equivalent.

3   http://www.harzing.com/pop.htm

4 It should be noted that also PoP lacks an API.

2. *Matching*: The list of retrieved GS entries will be parsed so as to identify the one(s) corresponding to the input publication. Given that an exact comparison for title matching may fail in many cases due to spelling errors, special characters, title extensions, and the general shortcomings in GS's parsing processes (see [TGM11] for a discussion), it was decided to identify the relevant GS entries based on approximate string matching by employing a string distance function with a given threshold. In addition, the year of publication would be matched if available[5]. This will result in a list of matching GS entries that will be fed to the next step.

3. *Merging*: The list of matching GS entries will be parsed so as to obtain their citations. When there is more than one matching GS entry, the union of the citations of all entries will be considered; duplicate entries will be identified by applying the strategy outlined in the previous step.

The same process will also be applied for collecting the MAS citation data with the difference that its API will be employed.

It should be noted that many researchers question whether purely computational approaches can ever generate reliable bibliographic and citation databases in a fully automatic manner without some human intervention to manually clean up and check the data [B08, B11, J06, J08, J11]. [TGM11] discusses in detail the shortcomings of such citation data sources (e.g., GS and MAS) which mainly stem from their parsing processes. These deficiencies need to be taken into account and addressed with manual data cleaning when possible. Our experience indicates that at least the second step of the above process (Matching) benefits enormously from human intervention and therefore it was decided that a manual validation step will be performed between the second and third steps of the above process.

Following the development of the above methodology, a detailed search through the bibliography on citation analysis led to the location of Professor Erhard Rahm's team in the University of Leipzig that has developed a tool for performing online citation analysis of computer science research[6]. For any set of DBLP publications, their Online Citation Service (OCS) system [TAR07] retrieves and integrates citation data on demand from four different data sources: GS, MAS, ACM Digital Library, and Citeseer. A set of search query generators is provided to efficiently retrieve relevant citation data and to iteratively refine search results for improved data quality. Given the close relation between their research and our goals, it was decided to contact them so as to establish a collaboration among the three research groups.

2.3 Data analysis

Dr. Tsikrika presented the types of citation data analyses that were performed in the preliminary study [TGM11] and it was decided that the analysis should be performed along the same axes. In addition, Dr. Larsen indicated the necessity of defining a baseline against which to compare the results. There is no straightforward answer in determining such as baseline given the interdisciplinary nature of evaluation campaigns such as ImageCLEF (and CLEF as whole) and the significant differences in the publishing and citing norms and practices among the different disciplines[7]. For instance, ImageCLEF focusses on the field of visual media analysis, indexing, classification, and retrieval, and to this end it develops evaluation tasks in various domains, including medical image annotation and retrieval, general image annotation and retrieval from historical archives, news photographic collections, and Wikipedia pages, robot vision, and plant identification. As a result, ImageCLEF participants originate from a number of different research communities, including (visual) information retrieval, cross–lingual information retrieval, computer vision and pattern recognition, medical informatics, and human-computer interaction, and thus their publications can be found in completely disparate "worlds". Given the differences in the publishing and citing practices between e.g., the disciplines of computer science and medicine, it is not trivial to define a baseline against which to compare the results of this ImageCLEF analysis as a whole. One solution that was proposed by Dr. Larsen would be to perform the analysis on the (relatively homogenous) task level. The publications and citations forming the baseline would then correspond to those in the related fields, e.g., the computer vision and pattern recognition field for the photo annotation task.

---

5   GS is not always able to correctly identify the publication year of an item [P06].

6   http://dbs.uni-leipzig.de/research/projects/citation_analysis

7   An interesting discussion on this can be found at: https://wiki.oulu.fi/display/tor/1.3.1.7+Evaluation+of+disciplines+and+research+fields.

3   Description of the main results obtained

The main outcome of this visit was the development of a methodology described in the previous section for assessing the scholarly impact of evaluation campaigns (and other research activities of a similar nature). This methodology is currently being applied in order to perform a scholarly impact analysis for ImageCLEF, and thus extend the preliminary study presented in [TGM11]. Our eventual goal is to assess the impact of the whole of CLEF.

Below you can find the main outcomes for each of the three main steps in the process regarding the scholarly impact assessment of ImageCLEF that is currently being conducted.

1. Publication data collection

    •   An initial "clean" (i.e., manually validated) set of publications *D* associated with ImageCLEF has been constructed. This consists of the 249 *Proceedings* and *Resources* publications used in the preliminary study [TGM11], the approximately 200 *WN* publications, and approximately 80 *Resources* and *Other* publications manually collected in collaboration with the ImageCLEF participants and organisers, i.e., around 550 publications in total.

    •   This dataset will be enlarged by applying the methodology presented in Section 2.1. A search using "imageclef" as the query in GS retrieves 1,660 articles, indicating that there is scope for considering many more publications as candidates for inclusion in the dataset to be analysed.

2. Citation data collection

    •   Google Scholar and Microsoft Academic Search have been selected as the citation data sources. Given that MAS has only recently been introduced and therefore has not yet achieved the coverage of GS, we have performed a preliminary comparison (see table below) on the number of citations each of these data sources finds for the 10 most cited ImageCLEF papers (as these were identified in [TGM11]). The comparison was first performed in October 2011 where the coverage of the two data sources is markedly different. A repeat of this comparison in February 2012 illustrates the significant improvements in MAS's coverage; for the top-7 publications the increase in the number of MAS citations in higher than the increase in the number of GS citations which indicates that this increase is not only due to the passage of time, but also to the improvements in the coverage. Our aim is to perform this comparison in regular intervals (e.g., once a month) so as to gauge the MAS's potential as a comprehensive citation data source.

ImageCLEF publication ID	October 2011		February 2012	
	GS citations	MAS citations	GS citations	MAS citations
1	49	11	56	33
2	44	21	54	29
3	40	3	42	27
4	32	7	37	18
5	32	5	66	55
6	18	0	25	9
7	16	5	21	13
8	14	0	21	3
9	13	0	13	1
10	10	1	21	3

    •   A tool for collecting GS and MAS citation data for any given set of publications will be developed. A collaboration has been established with Professor Erhard Rahm's team in the University of Leipzig so as to access their Online Citation Service (OCS) system. As a first step, we are currently in the process of providing them with the publication dataset used in [TGM11], together with the additional publications that have been added recently, so as to

acquire knowledge of their system and evaluate its performance by comparing and contrasting them with the results of our preliminary analysis. Depending on the results, the next step would be to consider the integration of their tool to our system through a web service.

3. Data analysis

- This step will be performed once the above steps have been completed.

4   Future collaboration with host institution (if applicable)

RSLIS recently (in March 2012) joined the PROMISE FP7 Network of Excellence where HES-SO is also participating. PROMISE aims to provide a virtual laboratory for conducting participative research and experimentation to carry out, advance and bring automation into the evaluation and benchmarking of such complex information systems, by facilitating management and offering access, curation, preservation, reuse, analysis, visualization, and mining of the collected experimental data. One of the PROMISE goals is to conduct an impact analysis for the CLEF initiative and it dedicates an entire task towards this goal; this task (led by RSLIS and involving HES-SO) provides the context within which the two teams will collaborate for the next 1,5 years.

5   Projected publications / articles resulting or to result from the grant

It is foreseen that an article on the extended study on the scholarly impact assessment of ImageCLEF (currently in preparation in collaboration with RLSIS and the University of Leipzig) will be submitted to the Journal of the American Society for Information Science and Technology before the end of 2012. Additional publications on methodological issues and on the eventual scholarly impact assessment of the whole of CLEF will also be considered.

6   Other comments (if any)

During her visit, Dr. Tsikrika was invited to present her research at two different institutions: the Royal School of Library and Information Science, Copenhagen, Denmark (i.e., the Host, see Section 6.1) and the University of Copenhagen, Denmark (see Section 6.2).

6.1 Invited talk at the Royal School of Library and Information Science, Copenhagen, Denmark (February 23, 2012)

Title: ImageCLEF evaluation activities

Abstract: ImageCLEF (http://www.imageclef.org/), the cross-language image retrieval annual evaluation campaign, was introduced in 2003 as part of the Cross Language Evaluation Forum (CLEF). Motivated by the need to support multilingual users from a global community accessing the ever growing body of visual information, the main goal of ImageCLEF is to support the advancement of the field of visual media analysis, indexing, classification, and retrieval, by (i) developing the necessary infrastructure for the evaluation of visual information retrieval systems operating in both monolingual and cross-language contexts, (ii) providing reliable and reusable resources for such benchmarking purposes, and (iii) promoting the exchange of ideas.

In this talk, Dr Tsikrika presented (i) an overview of the ImageCLEF evaluation tasks that benchmark the annotation and retrieval of diverse images such as general photographic and medical images, as well as domain-specific tasks such as plant identification and robot vision, (ii) the resources that have been built in the context of these evaluation activities and a study on their reliability and reusability, and (iii) a preliminary investigation on the scholarly impact of ImageCLEF indicated by the substantial numbers of its publications and their received citations.

URL: http://iva.dk/omiva/nyheder/insight/12-03-08/visuel-information-skal-synliggoeres/

6.2 Invited talk at the University of Copenhagen, Denmark (February 24, 2012)

Title: Exploring search log data

Abstract: Search engines unobtrusively record in their logs large amounts of a broad range of user-system search interactions. Search logs have been extensively analysed so as to gain an understanding of users' searching behaviour, while the clickthrough data recorded in them have been exploited in several Information Retrieval applications by being interpreted as users' implicit relevance feedback.

In the first part of this talk, Dr. Tsikrika presented research on search log analysis methods. Current methods describe the main features of the observed interactions in purely statistical terms without considering the semantics of the available information. We propose a semantic search log analysis method that enriches current approaches by exploiting the knowledge in a linked data cloud; particular focus has been placed on the analysis of users' behavioural patterns regarding query formulation and modification. A study of the search logs of the commercial picture portal of a European news agency that were analysed using the proposed method and the implications of the findings was also presented.

In the second part of the talk, Dr. Tsikrika presented research on the exploitation of clickthrough data for image annotation and retrieval applications and in particular (i) the use of clickthrough data for automatically generating labelled samples for training concept classifiers that perform image annotation, and (ii) the efficacy of topic modelling of clickthrough data in the image search domain.

URL: http://ccc.ku.dk/calendar/2012/exploring_search_log_data/

7   References

[B08] J. Bar-Ilan. Which h-index? A comparison of WoS, Scopus and Google Scholar. Scientometrics, 74(2):257–271, 2008.

[B11] D. Butler. Computing giants launch free science metrics. Nature 476 (18), 2011.

[H10] A.-W. Harzing. Citation analysis across disciplines: The impact of different data sources and citation metrics, 2010. http://www.harzing.com/data_metrics_comparison.htm

[IL07] P. Ingwersen and B. Larsen. Evaluation of strategic research programs: the case of Danish environmental research 1993–2002. Research Evaluation 16 (1): 47-57, 2007.

[J06] P. Jacsó. Deflated, inflated and phantom citation counts. Online Information Review, 30(3):297–309, 2006.

[J08] P. Jacsó. The pros and cons of computing the h-index using Google Scholar. Online Information Review, 32(3):437–452, 2008.

[J11] P. Jacsó. The pros and cons of Microsoft Academic Search from a bibliometric perspective. Online Information Review, 35(6): 983-997, 2011..

[MCDC10] H. Müller, P. Clough, T. Deselaers, and B. Caputo, editors. ImageCLEF: Experimental Evaluation in Visual Information Retrieval. Springer, 1st edition, 2010.

[NWL11] A. Névéol, W. J. Wilbur and Z. Lu. Extraction of data deposition statements from the literature: a method for automatically tracking research results. Bioinformatics 27(23): 3306-3312, 2011.

[RT05] E. Rahm and A. Thor. Citation analysis of database publications. SIGMOD Record, 34:48–53, December 2005.

[SLL08] J. W. Schneider, B. Larsen, and H. Lund. Report on the bibliometric analysis of the Danish ICT research from 2004-2007, Forsknings- og Innovationsstyrelsen, 2008. http://pure.iva.dk/en/publications/rapport-vedroerende-bibliometrisk-analyse-af-dansk-iktforskning-fra-20042007%2861dbb850-3e32-11de-8213-000ea68e967b%29.html

[TAR07] A. Thor, D. Aumueller, and E. Rahm. Data Integration Support for Mashups. In Proceedings of the Sixth International Workshop on Information Integration on the Web (IIWeb) 2007.

[TGM11] T. Tsikrika, A. G. Seco de Herrera, and H. Müller. Assessing the Scholarly Impact of ImageCLEF. In Proceedings of the Conference on Multilingual and Multimodal Information Access Evaluation (CLEF 2011), 19-22 September, Amsterdam, The Netherlands, 2011.