Assessing the Scholarly Impact of ImageCLEF

Theodora Tsikrika, Alba G. Seco de Herrera, and Henning Müller

University of Applied Sciences Western Switzerland (HES–SO)
Sierre, Switzerland
theodora.tsikrika@acm.org, alba.garcia@hevs.ch, henning.mueller@hevs.ch

Abstract. Systematic evaluation has an important place in information retrieval research starting with the Cranfield tests and currently with TREC (Text REtrieval Conference) and other evaluation campaigns. Such benchmarks are often mentioned to have an important impact in advancing a research field and making techniques comparable. Still, their exact impact is hard to measure. This paper aims at assessing the scholarly impact of the ImageCLEF image retrieval evaluation initiative. To this end, the papers in the proceedings published after each evaluation campaign and their citations are analysed using Scopus and Google Scholar. A significant impact of ImageCLEF could be shown through this bibliometric analysis. The differences between the employed analysis methods, each with its advantages and limitations, are also discussed.

1 Introduction

Evaluation campaigns in the field of information retrieval enable the reproducible and comparative evaluation of new approaches, algorithms, theories, and models, through the use of standardised resources and common evaluation methodologies within regular and systematic evaluation cycles. 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[1] (TREC), the Cross–Language Evaluation Forum[2] (CLEF), the INitiative for the Evaluation of XML retrieval[3] (INEX), and the NTCIR Evaluation of Information Access Technologies[4]. Similar evaluation exercises are also carried out in the field of visual information retrieval, with TREC Video Retrieval Evaluation[5] (TRECVid), PASCAL Visual Object Classes challenge[6], MediaEval[7], and ImageCLEF[8] being among the most prominent.

These evaluation campaigns are predominantly based on the Cranfield paradigm [2] of experimentally assessing the worth and validity of new ideas in a laboratory setting through the use of *test collections*. Although this evaluation model has met with some criticism (see [12] for a discussion), such organised

[1] http://trec.nist.gov/
[2] http://www.clef-campaign.org/
[3] http://www.inex.otago.ac.nz/
[4] http://ntcir.nii.ac.jp/
[5] http://trecvid.nist.gov/
[6] http://pascallin.ecs.soton.ac.uk/challenges/VOC/
[7] http://www.multimediaeval.org/
[8] http://www.imageclef.org/

benchmarking activities have been widely credited with contributing tremendously to the advancement of information retrieval by (i) providing access to infrastructure and evaluation resources that support researchers in the development of new approaches, and (ii) encouraging collaboration and interaction between researchers both from academia and industry. Their contribution to the field is mainly indicated by the research that would otherwise not have been possible, i.e., research that heavily relies on the use of resources they provide. It is then reasonable to consider that their success can be measured to some extent by the scientific and possibly the economic 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, and by additional indicators, such as filed patents, whereas its economic impact can be measured by the technology transfer efforts that result in commercial products and services. Other aspects, such as the scientific impact of the increased quality in evaluation methodologies or the economic impact of the time saved by researchers, who now reuse evaluation resources, rather than create them from scratch, are harder to assess. Recent investigations have reported on the scholarly impact of TRECVid [13] and on the economic impact of TREC [11]. Building on this work, this paper presents a preliminary study on assessing the scholarly impact of ImageCLEF, the cross–language image retrieval evaluation initiative that has been running as part of CLEF since 2003. To this end, it performs a citation analysis on a dataset of publications derived from ImageCLEF.

The rest of the paper is organised as follows: Section 2 provides a short introduction to ImageCLEF. Section 3 presents the bibliometric analysis method and tools, Section 4 the dataset of ImageCLEF publications considered, while Section 5 reports on the results of this analysis. Section 6 concludes this work.

2 The ImageCLEF Evaluation Campaign

ImageCLEF, the cross–language image retrieval annual evaluation campaign, was introduced in 2003 as part of CLEF and forms a natural extension to other CLEF tracks given the language neutrality of visual media. Motivated by the need to support multilingual users from a global community accessing the ever growing body of visual information, the main aims of ImageCLEF are: (i) to develop the necessary infrastructure for the evaluation of visual information retrieval systems operating in both monolingual and cross–language contexts, (ii) to provide reusable resources for such benchmarking purposes, and (iii) to promote the exchange of ideas towards the further advancement of the field of visual media analysis, indexing, classification, and retrieval.

To meet these objectives a number of tasks have been organised by ImageCLEF within two main domains: (i) medical image retrieval, and (ii) general (non–medical) image retrieval from historical archives, news photographic collections, and Wikipedia pages. These tasks can be broadly categorised as follows:

– *Ad hoc retrieval.* This simulates a classic document retrieval task: given a statement describing a user's information need, find as many relevant doc-

uments as possible and rank the results by relevance. In the case of cross–lingual retrieval, the language of the query is different from the language of the metadata used to describe the image. Ad hoc tasks have run since 2004 for medical retrieval and since 2003 for non–medical retrieval scenarios.
– *Object and concept recognition.* Although ad hoc retrieval is a core image retrieval task, a common precursor is to identify whether certain objects or concepts from a pre–defined set of classes are contained in an image (object class recognition), assign textual labels or descriptions to an image (automatic image annotation) or classify images into one or many classes (automatic image classification). Such tasks, including a medical image annotation and a robot vision task, have run since 2005.
– *Interactive image retrieval.* Since 2003, a user–centred task was run as a part of ImageCLEF and eventually followed by the interactive CLEF (iCLEF) track in 2005. Interaction in image retrieval can be studied with respect to how effectively the system supports users with query formulation, translation (for cross–lingual IR), document selection and examination.

Table 1 summarises the ImageCLEF tasks that ran between 2003 and 2010 and shows the number of participants for each task along with the distinct number of participants in each year. The number of participants and tasks offered by ImageCLEF has continued to grow steadily throughout the years, from four participants and one task in 2003, reaching its peak in 2009 with 65 participants and seven tasks. The number of participants, i.e., research groups that officially submit their results, is typically much smaller than the number of groups that register and gain access to the data; e.g., in 2010, 112 groups registered, but only 47 submitted results. Given its multi–disciplinary nature, 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.

Further information can be found in the ImageCLEF book [9] describing the formation, growth, resources, tasks, and achievements of ImageCLEF.

Table 1. Participation in the ImageCLEF tasks and number of participants by year.

Task	2003	2004	2005	2006	2007	2008	2009	2010
General images								
Photographic retrieval	4	12	11	12	20	24	19	–
Interactive image retrieval	1	2	*2*	*3*	–	*6*	*6*	–
Object and concept recognition				4	7	11	19	17
Wikipedia image retrieval						12	8	13
Robot vision							7	7
Medical images								
Medical image retrieval		12	13	12	13	15	17	16
Medical image annotation			12	12	10	6	7	–
Total (distinct)	4	17	24	30	35	45	65	47

3 Bibliometric Analysis Method

Bibliometric studies provide a quantitative and qualitative indication of the scholarly impact of research by examining the number of scholarly publications derived from it and the number of citations these publications receive. The most comprehensive sources for publication and in particular for citation data are: (i) the Thomson Reuters Web of Science[9] (generally known as ISI Web of Science or *ISI*), established by Eugene Garfield in the 1960s, (ii) *Scopus*[10], introduced by Elsevier in 2004, and (iii) the freely available *Google Scholar*[11], developed by Google in 2004. In addition to publication and citation data, ISI and Scopus also provide citation analysis tools to calculate various metrics of scholarly impact, such as the h–index [5], a robust metric of scientific research output that has a value h for a dataset of N_p publications, if h of them have at least h citations each, and the remaining $(N_p - h)$ publications have no more than h citations each. Google Scholar on the other hand is simply a data source and does not have such capabilities; citation analysis using its data can though be performed by the Publish or Perish[12] (PoP) system, a software wrapper for Google Scholar.

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, but only from 1996 onwards. Google Scholar, 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.

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 [1, 4]. For computer science, 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 [10, 1], given that its coverage of conference proceedings is only very partial, and thus disadvantages the impact of publications. For example, a recent study examining the effect of using different data sources for citation analysis across different disciplines [4] found that for a particular case of

[9] `http://apps.isiknowledge.com/`
[10] `http://www.scopus.com/`
[11] `http://scholar.google.com/`
[12] `http://www.harzing.com/pop.htm`

an established computer science academic, Scopus found 62% more publications and 43% more citations than ISI. Scopus' broader coverage can though be hindered by its lack of coverage before 1996, but this is not a problem in our case since the ImageCLEF evaluation campaign started in 2003. Google Scholar offers an even wider coverage than Scopus and thus further benefits citation analyses performed for the computer science field [10, 4]. As a result, this study employs both Scopus and Google Scholar (in particular its PoP wrapper) for assessing the scholarly impact of ImageCLEF. This allows us to also explore a further goal: to compare and contrast these two data sources in the context of such an analysis. Scopus and Google Scholar were also employed in the examination of the TRECVID scholarly impact [13], where emphasis was though mostly given on the Google Scholar data.

It should be noted that the reliability of Google Scholar as a data source for bibliometric studies is being received with mixed feelings [1], and some outright scepticism [7, 8]. This is due to its widely reported shortcomings [10, 7, 8, 1], which mainly stem from its parsing processes. In particular, Google Scholar frequently has several entries for the same publication, e.g., due to misspellings or incorrectly identified years, and therefore may deflate its citation count [10, 7]. This though can be rectified through PoP which allows for the manual merging of entries deemed to be equivalent. Inversely, Google Scholar may also inflate the citation count of some publications, since it may group together citations of different papers, e.g., the conference and journal version of a paper with the same or similar title or its pre–print and journal versions [10, 7]. Furthermore, Google Scholar is not always able to correctly identify the publication year of an item [7]. These deficiencies have been taken into account into our analysis and addressed with manual data cleaning when possible, but we should acknowledge that the validity of the citations in Google Scholar were not examined, since this is beyond the scope of this study. Next, we describe the collection and analysis of ImageCLEF publications and their citations using Scopus and PoP.

4 The Dataset of ImageCLEF Publications

CLEF's annual evaluation cycle culminates in a workshop where participants of all CLEF tracks (referred to as labs since 2010) present and discuss their findings with other researchers. This event is accompanied by the CLEF workshop proceedings, known as *working notes*, where research groups publish, separately for each track, notebook papers that describe the techniques used in their participation and results. In addition, the organisers of each track (and/or each task within each track) publish overview papers that present the evaluation resources used, summarise the approaches employed by the participating groups, and provide an analysis of the main evaluation results. The papers in the CLEF working notes are available online on the CLEF website and while they are not refereed, the vast majority of participants take the opportunity to publish there.

After the workshop, participants are invited to publish more detailed descriptions of their approaches and more in–depth analyses of the results of their participation, together with further experimentation, if possible, to the *CLEF proceedings*. These papers go through a reviewing process and the accepted ones, together with updated versions of the overview papers, are published in a volume of the Springer Lecture Notes in Computer Science series in the year following the workshop and the CLEF evaluation campaign. That means that the CLEF proceedings of the CLEF 2005 evaluation campaign were published in 2006. This publication scheme was followed until 2009; in 2010 the format of CLEF changed and the participants' and overview papers were only published in the CLEF working notes, i.e., there were no follow–up CLEF proceedings.

Moreover, CLEF participants may extend their work and publish in journals, conferences, and workshops. The same applies for research groups from academia and industry that, while not official participants of the CLEF activities, may decide at a later stage to use CLEF resources to evaluate their approaches. These *CLEF–derived* publications are a good indication of the impact of CLEF beyond the environment of the evaluation campaign. Furthermore, researchers directly involved with the development of CLEF evaluation resources and/or the coordination of tracks and tasks also publish elsewhere detailed descriptions of the applied methodologies, analyses of the reliability of the created resources, and best practices. These *CLEF resources* publications can be seen as complementary to the overview papers in the CLEF proceedings and working notes.

To assess the scholarly impact of ImageCLEF, bibliometric analysis can be applied to the dataset of publications that contains (i) the ImageCLEF–related publications in the *CLEF working notes* and (ii) in the *CLEF proceedings*, (iii) papers describing *ImageCLEF resources* (typically written by ImageCLEF organisers/coordinators), and (iv) *ImageCLEF–derived* publications where ImageCLEF datasets are employed for evaluating the research that is carried out. In this study, the dataset of publications that is analysed is formed as follows:

- *CLEF working notes*: Although publications in the CLEF working notes do attract citations (as discussed in the next section), given that Scopus does not index them, they are excluded from our analysis, so as to allow a "fair" comparison between the two citation data sources.
- *CLEF proceedings*: These publications are indexed by both Scopus and Google Scholar and therefore are included in our analysis. They were located by submitting a separate query for each of the CLEF proceedings published from 2004 to 2010 (and thus corresponding to the CLEF campaigns from 2003 to 2009, respectively). In Scopus, the query "SRCTITLE(lecture notes in computer science) AND VOLUME(*CLEF_proceedings_volume*) AND ALL(image OR photo OR imageclef* OR Flickr)" was entered in the Advanced Search. In PoP, 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. In both cases, the results were manually refined and cross–checked against the proceedings, so as to ensure that both the precision and recall of these results are 100%.
- *ImageCLEF resources*: Given that these publications are written by ImageCLEF organisers, they were located by searching by author name. The

results were manually refined by an expert in the field and added to the dataset of publications to be analysed.
- *ImageCLEF-derived publications*: Locating all publications that use ImageCLEF data is a hard task. One may assume that such papers would cite the overview article of the corresponding year of ImageCLEF, but often only the URL of the benchmark is mentioned, or that such papers are written by researchers having access to the data. Both such searches in Scopus and PoP require extensive manual data cleaning and the inclusion of such publications in the analysis is left as part of the next stage of our investigation.

Therefore, this preliminary study to assess the scholarly impact of ImageCLEF focusses on the analysis of the dataset of publications published between 2004 and 2010 and consisting of (i) ImageCLEF–related participants' and overview papers in the CLEF proceedings, and (ii) overview papers regarding ImageCLEF resources published elsewhere. The results are presented in Table 2 and are analysed in the next section.

5 Results for ImageCLEF Publications 2004–2010

The results of our study, presented in Table 2, show that there were a total of 195 ImageCLEF-related papers in the CLEF proceedings published between 2004 and 2010. Over the years, there is a steady increase in such ImageCLEF publications, in line with the continuous increase in participation and in the number of offered tasks (see Table 1). The coverage of publications regarding ImageCLEF resources varies greatly between Scopus and Google Scholar, with the former indexing a subset that contains only 57% of the publications indexed by the latter. These publications peak in 2010, which coincides with the year that ImageCLEF organised a benchmarking activity as a contest in the context of the International Conference for Pattern Recognition (ICPR). This event was accompanied by several overview papers describing and analysing the ImageCLEF resources used in the contest, published in the ICPR 2010 [6] and ICPR 2010 Contests [14] proceedings.

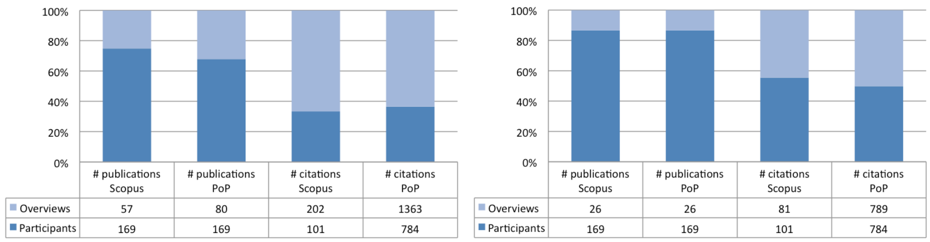
The number of citations varies greatly between Scopus and Google Scholar. For the publications in the CLEF proceedings, Google Scholar finds almost nine times more citations than Scopus. Apart from the wider coverage of Google Scholar, this is also partly due to its inability to distinguish in some cases publications with the same or similar title published in different venues, as is sometimes the case with papers published in the CLEF working notes and in the CLEF proceedings. Differentiating between the citations of two such versions of a CLEF paper requires extensive manual data cleaning that examines the list of references in the citing papers, a task which is beyond the scope of this study. Nevertheless, the inclusion of the citations to the CLEF working notes versions of some CLEF proceedings papers is considered acceptable in the context of this analysis, since they are still indicative of ImageCLEF's scholarly impact. When examining the distribution of citations over the years, Scopus indicates a variation in the number of citations, while Google Scholar shows a relative

Table 2. Overview of ImageCLEF publications 2004–2010 and their citations.

		CLEF proceedings			ImageCLEF resources			All		
	Year	papers	citations	h-index	papers	citations	h-index	papers	citations	h-index
Scopus	2004	5	13	2	4	31	3	9	44	4
	2005	20	50	4	–	–	–	20	50	4
	2006	25	24	3	3	28	1	28	52	3
	2007	27	25	2	6	29	2	33	54	3
	2008	29	18	3	5	22	2	34	40	3
	2009	45	14	2	2	4	1	47	18	2
	2010	44	38	4	11	7	2	55	45	4
	Total	195	182	6	31	121	5	226	303	9
Google Scholar	2004	5	65	3	5	105	4	10	170	6
	2005	20	210	8	5	47	4	25	257	10
	2006	25	247	7	8	144	5	33	391	9
	2007	27	259	7	10	76	4	37	335	9
	2008	29	249	7	7	73	5	36	322	9
	2009	45	284	7	7	53	4	52	337	9
	2010	44	259	7	12	76	6	56	335	10
	Total	195	1573	18	54	574	13	249	2147	22

stability from 2005 onwards. For publications regarding ImageCLEF resources, Google Scholar finds almost five times more citations than Scopus. These peak for papers published in 2006 and 2004, mainly due to three publications that describe the creation of test collections that were used extensively in ImageCLEF in the following years, and thus attracted many citations. Overall, Google Scholar indicates that the total number of citations over all 249 publications in the considered dataset are 2,147, resulting in 8.62 average cites per paper. This is comparable to the findings of the study on the scholarly impact of TRECVid [13], with the difference that they consider a much larger dataset of publications that also includes all TREC–derived papers.

Next, we analyse the distribution of citations over different types of papers starting with a comparison of the participants' papers in the CLEF proceedings with overviews describing ImageCLEF resources published both in the CLEF proceedings and elsewhere. Figure 1(a) compares the relative number of papers with the relative citation frequency for these publication types. While participants' papers account for a substantial share of the publications, namely 74.8% for Scopus and 67.9% for Google Scholar, they receive around 35% of the citations. Even when considering only the CLEF proceedings, i.e., when excluding the ImageCLEF resources papers published elsewhere so as to limit the bias towards overview papers that comes from including this dataset in the analysis, Figure 1(b) indicates that while participants' publications constitute 86.7% of the total, they attract around 50% of the citations. These results indicate the significant impact of the ImageCLEF overview papers.



(a) All (b) CLEF proceedings

Fig. 1. Relative impact of ImageCLEF publication types.

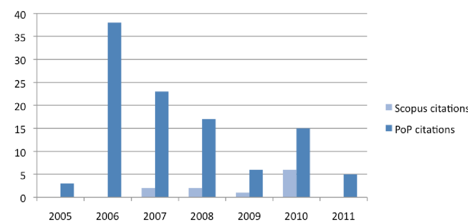


Fig. 2. Citations trends for the 2005 overview paper [3].

As an example, Figure 2 shows the evolution of the citations for the 2005 overview paper [3]. This is the last paper describing both medical and general tasks in a single overview, and as such it has been cited often. It shows a peak in the year after the competition and then slowly decreases with a half life of approximately three years in Google Scholar. In Scopus, the peak appears later and in general the number of citations remains almost stable over the years.

Next, the impact of publications in the two domains studied in ImageCLEF, medical and general images, is investigated. Figure 3 compares the relative number of publications with the citation frequency for the domains. It should be noted that some publications examine both domains at once, e.g., participants' papers presenting their approaches in ImageCLEF tasks that represent both domains or overview papers reporting on all tasks in a year. Therefore, the sum of publications (citations) in Figure 3 is not equal to the total listed in Table 2. Overall, the publications in the medical domain appear to have a slightly higher impact. To gain further insights, Figure 4 drills down from the summary data into the time dimension. At first, publications relating to the general domain dominate, with those relating to the medical domain increasing as the corresponding tasks establish themselves in the middle of the time period, while more recently there is again a shift towards the general domain. Scopus indicates that the impact of ImageCLEF publications that are related to the medical domain is particularly significant between 2006 and 2008. This is mostly due to number of overview papers regarding the medical image annotation task published both in the CLEF proceedings and elsewhere, and also because Scopus does not index

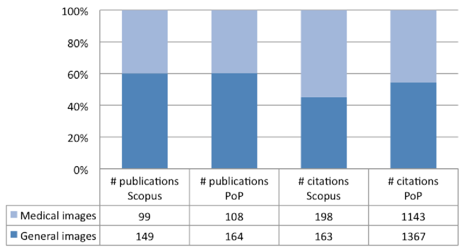
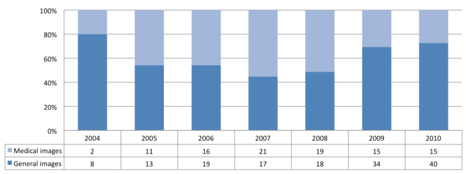
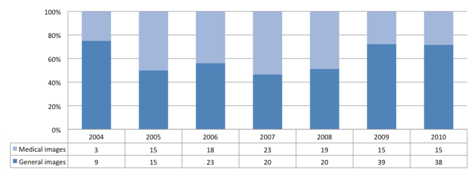


Fig. 3. Relative impact of ImageCLEF publications in the two domains.

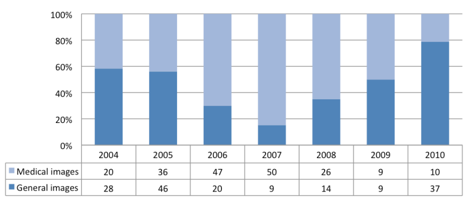


Scopus

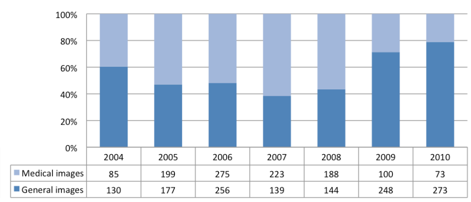


PoP

(a) publications



Scopus



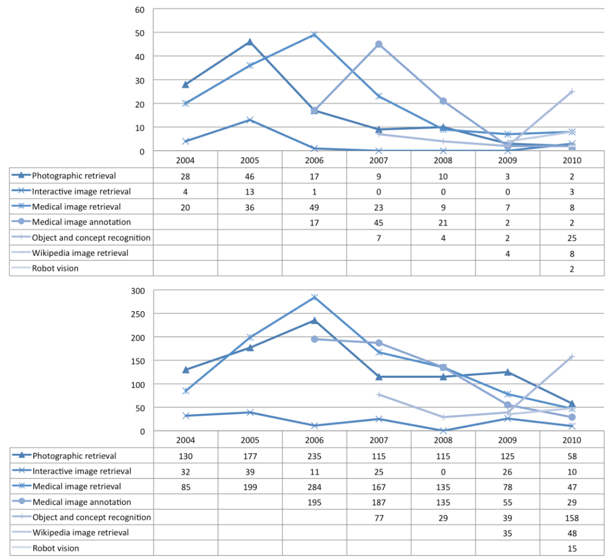
PoP

(b) citations

Fig. 4. Relative impact of ImageCLEF publications in the two domains over the years.

some of the ImageCLEF publications regarding general images that are found by Google Scholar. For Google Scholar, on the other hand, the distribution of citations appears to be mirroring that of the publications in the two domains.

Finally, Figure 5 depicts the distribution of citations for each of the ImageCLEF tasks (listed in Table 2) over the years. Similarly to above, a publication may cover more than one task. For all tasks, there is a peak in their second or third year of operation, followed by a decline. The exception is the object and concept recognition task, which attracts significant interest in its fourth year when it is renamed as photo annotation task and employs a new collection consisting of Flickr images and new evaluation methodologies. These novel aspects of the task result not only in increased participation (see Table 2), but also strengthen its impact. Overall, the photographic retrieval, the medical image retrieval, and the medical image annotation tasks have had the greatest impact.



	2004	2005	2006	2007	2008	2009	2010
Photographic retrieval	28	46	17	9	10	3	2
Interactive image retrieval	4	13	1	0	0	0	3
Medical image retrieval	20	36	49	23	9	7	8
Medical image annotation			17	45	21	2	2
Object and concept recognition				7	4	2	25
Wikipedia image retrieval						4	8
Robot vision							2

	2004	2005	2006	2007	2008	2009	2010
Photographic retrieval	130	177	235	115	115	125	58
Interactive image retrieval	32	39	11	25	0	26	10
Medical image retrieval	85	199	284	167	135	78	47
Medical image annotation			195	187	135	55	29
Object and concept recognition				77	29	39	158
Wikipedia image retrieval						35	48
Robot vision							15

Fig. 5. Citation trends per ImageCLEF task, Scopus (top) and PoP (bottom).

6 Conclusions

This paper aims at analysing the scholarly impact of the ImageCLEF image retrieval evaluation campaign. Both Scopus and Google Scholar are used to obtain the number of papers published in the course of ImageCLEF and their citations. This preliminary analysis concentrates on the CLEF post–workshop proceedings, as the CLEF working notes are not indexed by Scopus, and therefore a fair comparison between Scopus and Google Scholar, one of the goals of this study, would not have been possible. A few additional papers written by the organisers about the main workshop outcomes are added. A total of 249 publications were analysed obtaining 2,147 citations in Google Scholar and 303 in Scopus.

A comparison of Google Scholar and Scopus shows that both systems have advantages and limitations. Whereas Scopus is incomplete and misses many conference/workshop papers, the quality of its citation data is high. On the other hand Google Scholar is more complete, but contains errors such as combining publications with similar titles or having two entries for some publications.

An impact analysis over time shows that the half life of citations is around three years for the overview papers. The analysis also shows that tasks usually take a year to attract a larger number of participants but impact and participation usually drop after three years unless the task or the collection changes.

This analysis is only an intermediate step and it seems necessary to extend it to include not only the CLEF proceedings, but also the working notes and derived work. With the proceedings covering almost 230 papers and the non–reviewed working notes a larger number, 500 articles have already been published in this context. Taking into account the derived work, over 1,000 articles can be expected to be based on ImageCLEF data. It is also important to assess the impact of all of CLEF that contains 4–10 tasks and has run for 11 years, already.

This preliminary analysis shows ImageCLEF's significant scholarly impact through the substantial numbers of its publications and their received citations. ImageCLEF data has been used by over 200 research groups, many techniques have been compared during its campaigns, while its influence through imposing a solid evaluation methodology and through use of its resources goes even further.

7 Acknowledgements

The work was partially supported by the EU in the context of Promise (258191), Chorus+ (249008), and Khresmoi (257528) FP7 projects.

References

1. J. Bar-Ilan. Which h-index? A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2):257–271, 2008.
2. C. W. Cleverdon. The evaluation of systems used in information retrieval. In *Proceedings of the International Conference on Scientific Information (Vol. 1)*, pages 687–698. National Academy of Sciences, National Research Council, 1959.
3. P. Clough, H. Müller, T. Deselaers, M. Grubinger, T. M. Lehmann, J. Jensen, and W. Hersh. The CLEF 2005 cross–language image retrieval track. In *Proceedings of the 6th CLEF Workshop*, volume 4022 of *LNCS*, pages 535–557. Springer, 2006.
4. A.-W. Harzing. Citation analysis across disciplines: The impact of different data sources and citation metrics, 2010. Retrieved from `http://www.harzing.com/data_metrics_comparison.htm`.
5. J. E. Hirsch. An index to quantify an individuals scientific research output. *Proceedings of the National Academy of Sciences (PNAS)*, 102(46):16569–16572, 2005.
6. *Proceedings of the 20th International Conference on Pattern Recognition (ICPR 2010), Instanbul, Turkey.* IEEE Computer Society, 2010.
7. P. Jacsó. Deflated, inflated and phantom citation counts. *Online Information Review*, 30(3):297–309, 2006.
8. P. Jacsó. The pros and cons of computing the h-index using Google Scholar. *Online Information Review*, 32(3):437–452, 2008.
9. H. Müller, P. Clough, T. Deselaers, and B. Caputo, editors. *ImageCLEF: Experimental Evaluation in Visual Information Retrieval.* Springer, 1st edition, 2010.
10. E. Rahm and A. Thor. Citation analysis of database publications. *SIGMOD Record*, 34:48–53, December 2005.
11. B. R. Rowe, D. W. Wood, A. N. Link, and D. A. Simoni. Economic impact assessment of NIST's Text REtrieval Conference (TREC) Program. Technical Report Project Number 0211875, RTI International, 2010.
12. M. Sanderson. Test collection based evaluation of information retrieval systems. *Foundations and Trends in Information Retrieval*, 4:247–375, 2010.
13. C. V. Thornley, A. C. Johnson, A. F. Smeaton, and H. Lee. The scholarly impact of TRECVid (2003–2009). *JASIST*, 62(4):613–627, 2011.
14. D. Ünay, Z. Çataltepe, and S. Aksoy, editors. *Proceedings of the 20th International Conference on Recognizing Patterns in Signals, Speech, Images, and Videos, ICPR Contest Reports.* Springer, 2010.