The Scholarly Impact of CLEF (2000–2009)

Theodora Tsikrika[1], Birger Larsen[1], Henning Müller[2],
Stefan Endrullis[3], and Erhard Rahm[3]

[1] Royal School of Library & Information Science, University of Copenhagen, Denmark
[2] University of Applied Sciences Western Switzerland (HES–SO), Sierre, Switzerland
[3] University of Leipzig, Germany

Abstract. This paper assesses the scholarly impact of the CLEF evaluation campaign by performing a bibliometric analysis of the citations of the CLEF 2000–2009 proceedings publications collected through Scopus and Google Scholar. Our analysis indicates a significant impact of CLEF, particularly for its well-established Adhoc, ImageCLEF, and QA labs, and for the lab/task overview publications that attract considerable interest. Moreover, initial analysis indicates that the scholarly impact of ImageCLEF is comparable to that of TRECVid.

1 Introduction

The scholarly impact of *research activities* is commonly measured by their associated publications (i.e., the publications generated as a result of such activities) and the citations they receive. Existing work in bibliometrics and scientometrics has mainly focussed on assessing the scholarly impact of specific publication venues [5] (e.g., journals and conference proceedings) or of the research activities of individual authors [1], institutions, countries, or particular domains [2].

In the field of information retrieval, evaluation campaigns at the international level (e.g., TREC[1], CLEF[2], INEX[3], NTCIR[4], and FIRE[5]) constitute a research activity that has been widely credited with contributing tremendously to the advancement of the field. 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 retrieval systems. Given that their contribution to the feld 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 scholarly impact of the research they foster. Recent investigations have reported on the scholarly impact of TRECVid[6] [7] and ImageCLEF[7] [8]. Building on this

[1] Text REtrieval Conference (`http://trec.nist.gov/`)
[2] Cross–Language Evaluation Forum (`http://www.clef-initiative.eu/`)
[3] INitiative for the Evaluation of XML retrieval (`http://www.inex.otago.ac.nz/`)
[4] NTCIR Evaluation of Information Access Technologies (`http://ntcir.nii.ac.jp/`)
[5] Forum for Information Retrieval Evaluation (`http://www.isical.ac.in/~clia/`)
[6] TREC Video Retrieval Evaluation (`http://trecvid.nist.gov/`)
[7] CLEF Image Retrieval Evaluation (`http://www.imageclef.org/`)

P. Forner et al. (Eds.): CLEF 2013, LNCS 8138, pp. 1–12, 2013.

work, this paper presents a preliminary study on assessing the scholarly impact of the first ten years of CLEF activities. To this end, it performs a citation analysis on a dataset of publications obtained from the CLEF proceedings.

2 The CLEF Evaluation Campaign

Evaluation campaigns 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. Motivated by the need to support users from a global community accessing the ever growing body of multilingual and multimodal information, the CLEF annual evaluation campaign, launched in 1997 as part of TREC, became an independent event in 2000 with the goal to promote research, innovation, and development of information access systems with an emphasis on multilingual and multimodal information. To this end, it provides an infrastructure for: (i) the comparative evaluation of multilingual and multimodal information access systems, (ii) the creation of reusable resources for such benchmarking purposes, (iii) the exploration of new evaluation methodologies and innovative ways of using experimental data, and (iv) the exchange of ideas.

CLEF is organised as a series of evaluation *labs* (referred to as *tracks* before 2010), each with a focus on a particular research area, ranging from the core cross-lingual adhoc retrieval (*Adhoc*) to multilingual question answering (*QA@CLEF*), cross-language image retrieval (*ImageCLEF*), and interactive retrieval (*iCLEF*). Some labs are in turn structured into tasks, each with even more focussed research objectives. In 2010, CLEF changed its format by accompanying its labs with a peer-reviewed conference. This paper focusses on the first ten years of CLEF and does not consider the changes that took place thereafter.

CLEF's annual evaluation cycle culminates in a workshop where participants of all labs present and discuss their findings with other researchers. This event is accompanied by the **CLEF working notes**, where research groups publish, separately for each lab and task, *participant* notebook papers that describe their techniques and results. In addition, the organisers of each lab (and/or each task) 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. Moreover, *evaluation* papers reflecting on evaluation issues, presenting other evaluation initiatives, or describing and analysing evaluation resources and experimental data may also be included. These (non-refereed) CLEF working notes papers are available online on the CLEF website.

From 2000 to 2009, participants were invited to publish after each workshop 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 went through a reviewing process and the accepted ones, together with updated versions of the overview papers, were published in a volume of the Springer Lecture Notes in Computer Science series in the year following the workshop and the CLEF evaluation campaign.

Moreover, CLEF participants and organisers 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.

3 Bibliometric Analysis Method

Bibliometric studies provide a quantitative and qualitative indication of the scholarly impact of research by examining the number of publications derived from it and the number of citations these publications receive. The most comprehensive citation data sources are: (i) *ISI* Web of Science, (ii) *Scopus*, and (iii) *Google Scholar*. ISI and Scopus also provide citation analysis tools to calculate various metrics of scholarly impact, such as the h–index [3]. Google Scholar, on the other hand, does not offer such capabilities for arbitrary publication sets; citation analysis using its data can though be performed by systems such as the *Online Citation Service* (OCS – `http://dbs.uni-leipzig.de/ocs/`) and *Publish or Perish* (PoP – `http://www.harzing.com/pop.htm`).

Each of these sources follows a different data collection policy that affects both the publications covered and the number of citations found. Differences in their coverage can enormously affect the assessment of scholarly impact metrics; the degree to which this happens varies among disciplines [1,2]. For computer science, where publications in peer–reviewed conference proceedings are highly valued and cited in their own right, ISI greatly underestimates the number of citations found [5,1], given that its coverage of conference proceedings is very partial. Scopus offers broader coverage, which may though be hindered by its lack of coverage before 1996; this does not affect this study. Google Scholar offers an even wider coverage and thus further benefits citation analyses performed for the computer science field [5,2]. As a result, this study employs both Scopus and Google Scholar (in particular its OCS and PoP wrappers) for assessing the scholarly impact of CLEF. This allows us to also explore a further goal: to compare and contrast these data sources in the context of such an analysis.

Similarly to [8], the focus is on the CLEF proceedings; analysis of the CLEF working notes and CLEF-derived publications is left as future work. The CLEF 2000–2009 proceedings contain 873 publications. These were obtained through DBLP and were semi-automatically annotated with their *type* (i.e., evaluation, participant, or overview) and the *lab(s)* and/or *tasks(s)* they refer to.

Their citations were obtained as follows in an 24-hour period in April 2013. In Scopus, the query "SRCTITLE(lecture notes in computer science) AND VOLUME(*proceedings_volume*)" was entered in the Advanced Search separately for each year and the results were manually cross–checked against the publication list. In OCS, the list of publications was directly uploaded to the system which matched them to one or more Google Scholar entries. The result list consisting of tuples of the form <*input_publication*, *Google_Scholar_match*, *number_of_citations*> was manually refined so as to remove false positive matches.

Furthermore, the citations (if any) of publications for which OCS did not find a match were manually added to the list. In PoP, the proceedings title was used in the Publication field and the proceedings publication year in the Year field. The results were also manually refined by removing false positive matches, merging entries deemed equivalent, and adding the citations of unmatched publications.

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 [4], due to its widely reported shortcomings [5,4,1]. 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 citation counts [5,4]. OCS rectifies this through multiple matching and PoP through support for manual merging. Inversely, Google Scholar may also inflate citation counts by grouping together citations of different papers, e.g., the journal and conference version of a paper with the same or similar titles [5,4]. Furthermore, Google Scholar is not always able to correctly identify the publication year of an item [4]. These deficiencies have been taken into account and addressed with manual data cleaning when possible, but we should acknowledge that examining the validity of citations in Google Scholar is beyond the scope of this study.

4 Results of the Bibliometric Analysis

The results of the bibliometric analysis of the citation data found by the three sources for the 873 CLEF proceedings publications are presented in Table 1. Over the years, there is a steady increase in the number of publications, in line with the continuous increase in the number of offered labs (with the exception of 2007). The coverage of publications varies significantly between Scopus and Google Scholar, with the former indexing a subset that does not include the entire 2000 and 2001 CLEF proceedings and another four individual publications, and thus contains 92% of all publications, while the latter does not index 22 (0.02%) of all publications. Table 2 indicates that Spain is the country that has produced the most CLEF proceedings publications, with five of its institutions and four of its authors being among the top 10 most prolific. Although the statistics in Table 2 are obtained from Scopus, and therefore cover only the years 2002–2009, they can still be considered representative of the whole dataset since they describe over 90% of all publications; OCS and PoP do not readily support such analysis.

The number of citations varies greatly between Scopus and Google Scholar, with the latter finding around ten times more citations than Scopus. Overall, the total number of citations over the 873 CLEF proceedings publications are 9,137 and 8,878 as found by OCS and PoP, respectively, resulting in 10.47 and 10.17 average cites per paper, respectively, while Scopus only finds 905 citations.

The differences between these data sources are investigated further by examining the correlations of the citations they find. Scopus' low coverage does not allow for meaningful comparisons to the other two sources and therefore our investigation focusses on the differences between OCS and PoP. Since both rely on Google Scholar, their differences are not substantial. Figure 1(a) shows a strong

Table 1. The citations, average number of citations per publication, and h-index of the CLEF proceedings publications as found by the three sources

	# labs	# publ.	OCS			PoP			Scopus		
			# cit.	avg.	h-index	# cit.	avg.	h-index	# cit.	avg.	h-index
2000	3	27	501	18.56	15	507	18.78	15	-	-	-
2001	2	37	904	24.43	17	901	24.35	17	-	-	-
2002	4	44	636	14.45	14	634	14.41	14	74	1.68	4
2003	6	65	787	12.11	15	776	11.94	15	87	1.34	5
2004	6	81	989	12.21	17	942	11.63	16	137	1.69	5
2005	8	112	1231	10.99	18	1207	10.78	17	133	1.19	5
2006	8	127	1278	10.06	18	1250	9.84	18	133	1.05	5
2007	7	116	1028	8.86	16	902	7.78	15	119	1.03	5
2008	10	131	1002	7.65	16	989	7.55	16	78	0.60	3
2009	10	133	781	5.87	12	770	5.79	12	144	1.08	5
Total	14	873	9,137	10.47	41	8,878	10.17	41	905	1.04	10

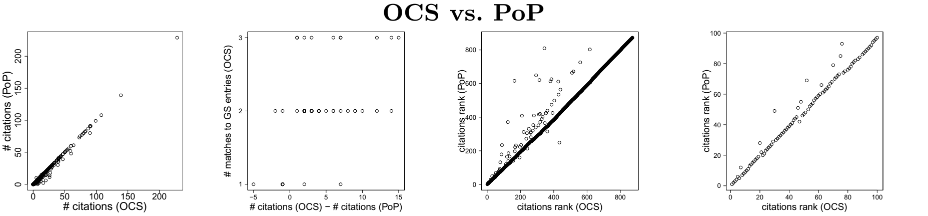
Table 2. Top 10 countries, affiliations, and authors of the CLEF 2002–2009 proceedings publications as found by Scopus

Country		Affiliation		Author	
Spain	178	Universidad de Alicante	44	Jones G.J.F.	29
Germany	105	UNED	33	Mandl T.	25
United States	93	Dublin City University	30	Llopis F.	24
France	67	University of Amsterdam	29	de Rijke M.	24
United Kingdom	61	Universidad de Jaen	27	Garcia-Cumbreras M.A.	20
Italy	55	Universität Hildesheim	25	Urena-Lopez L.A.	20
Netherlands	54	Universidad Carlos III de Madrid	24	Clough P.	19
Switzerland	52	UC Berkeley	23	Penas A.	18
Ireland	41	Universidad Politecnica de Madrid	22	Rosso P.	18
Canada	25	University of Sheffield	21	Leveling J.	17

correlation between the number of citations OCS and PoP find for each publication, particularly for publications with high citation counts. This is further confirmed by Figures 1(c)–(d) that show the correlations between the rankings based on the citation counts over all publications and over the 100 most cited publications, respectively. Here, ties in the rankings are resolved using the titles, but similar results are obtained when using the authors' names. The overlap in publications ranked by both in the top $k = \{100, 200, 300, 400, 500\}$ is over 96%.

Overall, OCS finds 259 (3%) more citations than PoP. The difference for a single publication ranges from 1 to 15 citations, as illustrated in Figure 1(b). Small differences could be attributed to changes in the Google Scholar index that may have taken place during the time period that intervened between obtaining the citation data from each source. Larger differences could be attributed to the different policies adopted by OCS and PoP for matching each input publication to a Google Scholar entry. Figure 1(b) plots the differences in citation counts against the number of Google Scholar matches found by OCS; the higher the difference, the more likely that OCS found more matches. This indicates that OCS achieves a slightly higher recall, and therefore OCS data will be used for the analysis performed in the following sections, unless stated otherwise.

Finally, when examining the distributions over the years, OCS and PoP reach their peak in terms of number of citations and h-index values in 2006. The average number of citations per publication peaks much earlier though, indicating



(a) all (b) diff. vs. # matches (c) rank-based (all) (d) rank-based (top 100)

Fig. 1. Correlations between the citations found by the different sources

that the publications of the early CLEF years have on average much more impact than the more recent ones. This could be attributed to the longer time period afforded to these earlier publications for accumulating citations. Given though the current lack of access to the citing papers through the OCS and PoP systems, only a future analysis that will monitor changes in regular intervals (e.g., yearly) could provide further insights (see also Section 4.4).

4.1 Citation Distribution

Metrics such as the total number of citations and the average number of citations per publication do not allow us to gauge the impact of individual publications, given that scientific publications are typically cited to a variable extent and citation distributions across such publications are found to be highly skewed [6]. To determine the degree of citation skew and thus gain insights into the variability of the impact of particular publications, the distribution of citations into publication quartiles are examined for each year and overall.

Figure 2 indicates the relative cumulative citation count for each quartile of publications. The 25% of top cited publications account for 50 to 75% of all citations (72% on average), while the bottom 25% of publications merely attract 0.5–7.5% of all citations (1.5% on average). This citation skewness appears to be increasing over the years. For the first three years, the top 25% of publications account for less than 60% of all citations, for the next three years, for around 65% of all citations, while for the last four years, for close to 75% of all citations.

These results are corroborated by also measuring the skewness of the citation distribution using the *Gini coefficient*, a measure of statistical dispersion that reflects the inequality among values of a frequency distribution. The Gini coefficient corresponds to a nonnegative real number, with higher values indicating more diverse distributions; 0 indicates complete equality, and 1 total inequality. Its overall value of 0.63 in CLEF indicates the high degree of variability in the citations of individual publications, and this diversity is continuously increasing as indicated by the values of the Gini coefficient being below 0.5, around 0.55, and over 0.65 for the first three, next three, and final four years, respectively.

The exception to the above observations is the year 2001, which is more skewed compared to the other early CLEF years; its Gini coefficient is 0.61, while its

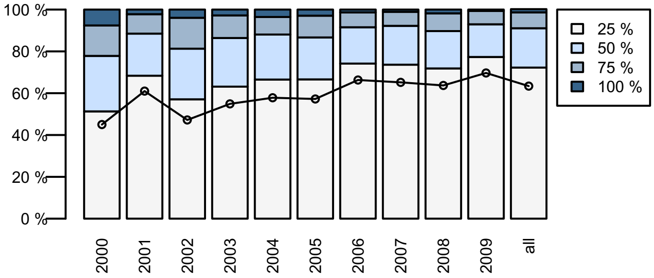


Fig. 2. The distributions of citations found by OCS (split by quarters) over the years and overall, and the Gini coefficient of these distributions plotted as a line

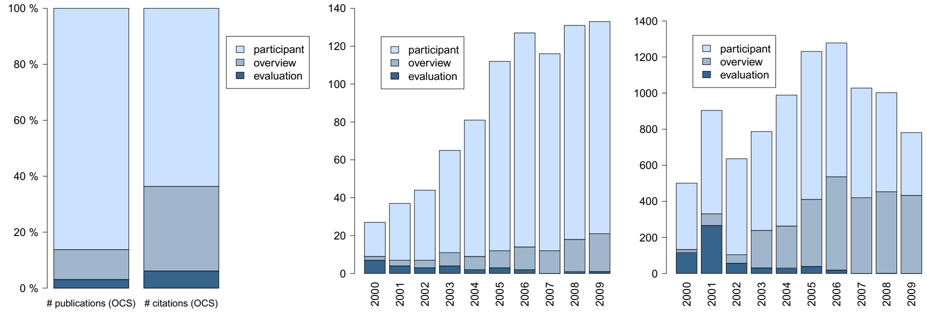
Table 3. Top 10 cited publications as found by OCS: their rank and number of citations by the three sources, and their author(s), title, year, and type (E = *evaluation*, O = *overview*, P = *participant*). Terms in *italics* denote abbreviations of original title terms.

OCS / PoP / Scopus						Author(s)	Title	Year	Type
rank			# citations						
1	1	-	228	229	-	Voorhees	The Philosophy of Information Retrieval Evaluation.	2001	E
2	2	2	139	139	17	Müller et al.	Overview of the ImageCLEFmed 2006 Medical Retrieval [...]	2006	O
3	3	5	108	108	12	Clough et al.	The CLEF 2005 Cross-Language Image Retrieval Track.	2005	O
4	4	1	99	99	17	Clough et al.	The CLEF 2004 Cross-Language Image Retrieval Track.	2004	O
5	6	290	91	91	4	Vallin et al.	Overview of the CLEF 2005 Multilingual *QA* Track.	2005	O
6	5	6	90	91	11	Chen	Cross-Language Retrieval Experiments at CLEF 2002.	2002	P
7	12	29	90	80	5	Grubinger et al.	Overview of the ImageCLEFphoto 2007 [...] Task.	2007	O
8	7	-	90	90	-	Monz & de Rijke	Shallow Morphological Analysis in Monolingual *IR* [...]	2001	P
9	8	14	87	87	7	Müller et al.	Overview of the CLEF 2009 Medical Image Retrieval Track.	2009	O
10	9	4	83	83	13	Magnini et al.	Overview of the CLEF 2004 Multilingual *QA* Track.	2004	O

top 25% publications account for almost 70% of all citations. This high degree of variability is due to the inclusion of two of the top 10 cited publications over all years, listed in Table 3, and in particular due to the domination of the most cited publication, a paper by Ellen Voorhees [9], which achieves around 65% more citations than the second most cited publication. The remaining top cited publications in Table 3 are more or less evenly spread across the years.

4.2 Citation Analysis of CLEF Publications Types

Figure 3(a) compares the relative number of publications of the three types (*evaluation*, *overview*, and *participant*) with their relative citation frequency. As also listed in the last column of Table 4, the participants' publications account for a substantial share of all publications, namely 86%, but only receive 64% of all citations. On the other hand, overview and evaluation publications receive three times or twice the percentage of citations compared to their publications' percentage. This indicates the significant impact of these two types; the significant impact of overview publications is further illustrated in Table 3 where 7 out of the 10 most cited publications are overviews, while the impact of evaluation publications can be attributed to a single publication, the Voorhees paper [9].



(a) all (b) # publications (OCS) (c) # citations (OCS)

Fig. 3. Relative impact of different types of CLEF proceedings publications

Table 4. Relative percentages of different types of CLEF proceedings publications and their citations over the years

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2000–2009
	% publications										
evaluation	25.93	10.81	6.82	6.15	2.47	2.68	1.57	0.00	0.76	0.75	0.03
overview	7.41	8.11	9.09	10.77	8.64	8.04	9.45	10.34	12.98	15.04	0.11
participant	66.67	81.08	84.09	83.08	88.89	89.29	88.98	89.66	86.26	84.21	0.86
	% citations										
evaluation	23.15	29.42	8.96	3.94	3.03	3.17	1.49	0.00	0.10	0.00	0.06
overview	3.39	7.19	7.39	26.43	23.56	30.22	40.45	40.86	45.11	55.44	0.30
participant	73.45	63.38	83.65	69.63	73.41	66.61	58.06	59.14	54.79	44.56	0.64

Figures 3(b)–(c) and Table 4 drill down from the summary data into the time dimension. During the early years, CLEF proceedings included several evaluation publications, many of them invited, which attracted a considerable number of citations, with the Voorhees [9] paper in 2001 being the most prominent example. More recently, such publications and consequently their citations have all but disappeared. The number of participants' publications has mostly followed a steady increase both in absolute and in relative terms, reaching almost 90% of all publications for some years. However, such publications manage to attract only between 44% and 74% of all citations, with the exception of 2002, where participants' publications received almost 84% of all citations. This is mostly due to a single participant's publication included among the 10 most cited publications (see Table 3). Finally, the impact of overview publications has significantly increased during the more recent years, where overviews constitute only 10 to 15% of all publications, but account for 40 to 55% of all citations.

4.3 Citation Analysis of CLEF Labs and Tasks

Table 5 presents the results of the citation analysis for the publications of the 14 labs and their tasks organised by CLEF during its first 10 years. Two more "pseudo–labs", *CLEF* and *Other* are also listed; these are used for classifying the

Table 5. CLEF labs and tasks in alphabetical order, the number of years they have run, their publications, citations, average number of citations per publication, and the type of the most cited publication (E = *evaluation*, O = *overview*, P = *participant*). The number of publications and citations over all tasks for a lab may not sum up to the total listed for *all tasks* for that lab, since a publication may refer to more than one task. Similarly for the number of publications and citations over all labs.

Lab	**Task**	**#years**	**# publications**	**# citations**	**average**	**most cited**
Adhoc	(*all tasks*)	10	237	2540	10.72	P
	Cross/Mono-lingual	8	188	2285	12.15	P
	Persian	2	11	97	8.82	O
	Robust	4	30	192	6.40	O
	TEL	2	19	150	7.89	O
CL-SR		6	29	208	7.17	O
CLEF		10	23	203	8.83	E
CLEF-IP		1	15	85	5.67	O
Domain-Specific		9	47	555	11.81	P
GeoCLEF		4	58	561	9.67	O
GRID@CLEF		1	3	8	2.67	O
iCLEF		9	41	378	9.22	O
ImageCLEF	(*all tasks*)	7	179	2018	11.27	O
	Interactive	1	2	4	2.00	P
	Medical Annotation	5	37	586	15.84	O
	Medical Retrieval	6	62	1002	16.16	O
	Photo Annotation	4	21	245	11.67	O
	Photo Retrieval	7	86	1002	11.65	O
	Robot Vision	1	6	23	3.83	O
	Wikipedia Retrieval	2	11	74	6.73	O
INFILE		2	8	5	0.62	O
LogCLEF		1	6	25	4.17	O
MorphoChallenge		3	20	247	12.35	P
Other		5	8	277	34.62	E
QA@CLEF	(*all tasks*)	7	173	2023	11.69	O
	AVE	3	25	274	10.96	O
	GikiCLEF	1	7	32	4.57	O
	QA	6	114	1489	13.06	O
	QAST	3	11	89	8.09	O
	ResPubliQA	1	10	95	9.50	O
	WiQA	1	7	52	7.43	O
VideoCLEF		2	14	79	5.64	O
WebCLEF		4	28	180	6.43	P
All		10	873	9,137	10.47	E

evaluation type publications not assigned to specific labs, but rather pertaining to evaluation issues related to CLEF or other evaluation campaigns, respectively.

Three labs, *Adhoc*, *ImageCLEF*, and *QA@CLEF*, clearly dominate in terms of publication and citation numbers; they account for 67% of all publications and for 72% of all citations. They also account for 9 of the 10 most cited publications in Table 3. The highest number of citations per publication is observed for the Other evaluation publications, which are highly skewed due to the presence of the Voorhees [9] paper. Excluding these from further consideration, the aforementioned three labs are among the top ranked ones, together with the *Domain–Specific* and *MorphoChallenge*. Overall, the *Medical Retrieval* and *Medical Annotation* ImageCLEF tasks have had the greatest impact among all labs and tasks, closely followed by the main *QA* task and the main *Cross/Monolingual* Adhoc task. This also indicates a bias towards older, most established labs and tasks. Finally, the most cited publication in each lab or task is in most cases its overview, further indicating the high impact of such publications.

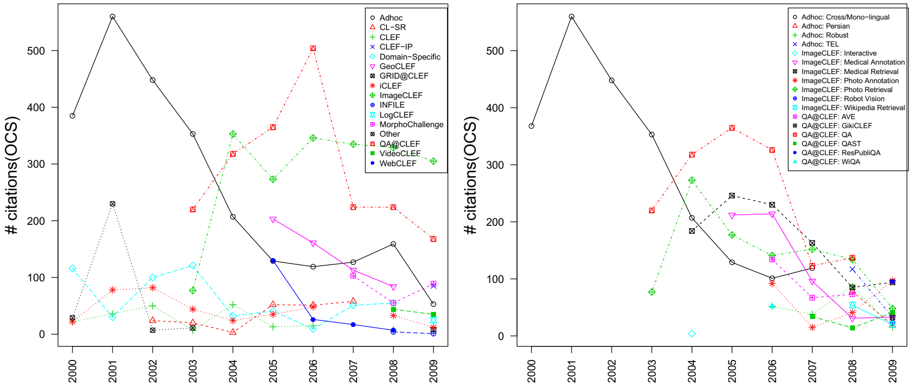


Fig. 4. The impact of CLEF labs (left) and tasks (right) over the years

Figure 4 depicts the number of citations for the CLEF labs and tasks over the years. Although it is difficult to identify trends over all labs and tasks, in many cases there appears to be a peak in their second or third year of operation, followed by a decline. Exceptions include the *Photo Annotation* ImageCLEF task, which attracted significant interest in its fourth year when it employed a new collection and adopted new evaluation methodologies, and also the *Cross–Language Speech Retrieval* (CL–SR) lab that increased its impact in 2005 following a move from broadcast news to conversational speech. Such novel aspects result in renewed interest in labs and tasks, and also appear to strengthen their impact.

4.4 Assessing the Impact of ImageCLEF in 2011 and in 2013

A previous study [8] assessed the scholarly impact of ImageCLEF by performing a bibliometric analysis of citation data collected in April 2011 through Scopus and PoP. Table 6 compares and contrasts the results of this earlier study with the results of this work using the same data sources two years later. The earlier study also took into account iCLEF publications that relied on ImageCLEF datasets or were otherwise closely related to ImageCLEF. However, the impact of these additional publications is negligible, since their citations account for less than 0.04% of all citations; these two results sets can be viewed as being comparable.

There is a considerable increase in the number of citations over these two years: 364 (+23%) more citations are found by PoP and 91 (+50%) by Scopus. For PoP, most citations are added to the 2004 and 2006 publications, while for Scopus to the 2007–2009 ones. Overall, the impact of ImageCLEF tasks appears to increase several years after they took place, however further analysis is needed to determine whether these citations originate from papers published over these two years, or from papers simply added to the sources' indexes during this time.

Table 6. Bibliometric analyses of the ImageCLEF publications published in the CLEF 2003–2009 proceedings performed in 2011 and in 2013 using Scopus and PoP

		#publications		**# citations**		**average**		**h-index**	
		2011	2013	2011	2013	2011	2013	2011	2013
Scopus	2003	5	5	13	14	2.60	2.80	2	3
	2004	20	20	50	64	2.50	3.20	4	5
	2005	25	22	24	30	0.96	1.36	3	3
	2006	27	23	25	38	0.93	1.65	2	3
	2007	29	29	18	34	0.62	1.17	3	3
	2008	45	40	14	34	0.31	0.85	2	3
	2009	44	40	38	59	0.86	1.48	4	5
	Total	195	179	182	273	0.93	1.53	6	7
PoP	2003	5	5	65	74	13.00	14.80	3	4
	2004	20	20	210	340	10.50	17.00	8	10
	2005	25	22	247	265	9.88	12.05	7	8
	2006	27	23	259	344	9.59	14.96	7	8
	2007	29	29	249	291	8.59	10.03	7	9
	2008	45	40	284	318	6.31	7.95	7	8
	2009	44	40	259	305	5.89	7.63	7	7
	Total	195	179	1,573	1,937	8.06	10.82	18	22

Table 7. Bibliometric analyses of all TRECVid (*TVa*) [7], TRECVid working notes (*TV*), CLEF proceedings (*C*), and ImageCLEF (*I*) publications using PoP

	#publications				**# citations**				**average**				**h-index**			
	TVa	TV	C	I	TVa	TV	C	I	TVa	TV	C	I	TVa	TV	C	I
2003	64	27	65	5	1,066	561	787	74	16.66	20.78	12.11	14.80	18	10	15	4
2004	158	29	81	20	2,124	423	989	340	13.44	14.59	12.21	17.00	24	11	17	10
2005	225	26	112	22	2,537	433	1231	265	11.28	16.65	10.99	12.05	28	8	18	8
2006	361	35	127	23	4,068	437	1278	344	11.27	12.49	10.06	14.96	30	11	18	8
2007	382	34	116	29	3,562	244	1028	291	8.97	7.18	8.86	10.03	28	6	16	9
2008	509	40	131	40	1,691	175	1002	318	3.32	4.37	7.65	7.95	16	10	16	8
2009	374	13	133	40	780	12	781	305	2.09	0.92	5.87	7.63	12	2	12	7
Total	2,073	205	765	179	15,828	2,285	7,096	1,937	7.63	11.21	9.28	10.82	52	25	38	22

4.5 Comparing to the Impact of other Evaluation Campaigns

Assessments of the scholarly impact of other evaluation campaigns have only been performed for TRECVid (2003–2009) [7], where a list containing both the *TRECVid working notes* and the *TRECVid–derived* publications was analysed. For comparability to the CLEF proceedings, we obtained the data used in [7] (`http://www.cdvp.dcu.ie/scholarly-impact/`) and manually identified the subset of the TRECVid working notes publications. Table 7 analyses these three sets (all TRECVid, TRECVid working notes, CLEF publications), and also the ImageCLEF publications, since this lab and TRECVid focus on similar domains.

Overall, there are about three times more TRECVid publications than CLEF proceedings ones, but receive on average less citations. It is difficult though to draw conclusions given the multidisciplinary nature of CLEF coupled with the different citation practices in different domains. The number of TRECVid working notes publications is close to that of ImageCLEF, with the former attracting a slightly higher number of citations, but not significantly so; both perform better than the larger sets. It appears that ImageCLEF is on par with TRECVid, taking also into account the fact that ImageCLEF was first established in 2003, while TRECVid was part of TREC already from 2001 and became an independent event in 2003. On the other hand, the TRECVid working notes publications list is rather incomplete (cf. [7]). Also, the data in [7] were collected earlier and thus it is likely that the TRECVid publications have attracted more citations over time. Further investigation is needed for reaching more reliable conclusions.

5 Conclusions

Measuring the impact of evaluation campaigns may prove useful for supporting research policy decisions by determining which aspects have been successful, and thus obtaining guidance for the development of improved evaluation methodologies and systems. This bibliometric analysis of the CLEF 2000–2009 proceedings has shown the considerable impact of CLEF during its first ten years in several diverse multi-disciplinary research fields. The high impact of the overview publications further indicates the significant interest in the created resources and the developed evaluation methodologies, typically described in such papers. It is necessary though to extend this analysis and include the working notes and all derived work. Finally, our analysis has highlighted the differences between the available citation analysis tools: Google Scholar provides a much wider coverage than Scopus, while OCS and PoP are in essence comparable, each with different querying facilities that might prove advantageous in different situations.

Acknowledgements. This work was partially supported by the ELIAS ESF Research Networking Programme, and by the Promise (258191) and Khresmoi (257528) FP7 projects.

References

1. Bar-Ilan, J.: Which h-index? A comparison of WoS, Scopus and Google Scholar. Scientometrics 74(2), 257–271 (2008)
2. Harzing, A.-W.: Citation analysis across disciplines: The impact of different data sources and citation metrics (2010), `http://www.harzing.com/data_metrics_comparison.htm` (retrieved)
3. Hirsch, J.E.: An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences (PNAS) 102(46), 16569–16572 (2005)
4. Jacsó, P.: Deflated, inflated and phantom citation counts. Online Information Review 30(3), 297–309 (2006)
5. Rahm, E., Thor, A.: Citation analysis of database publications. SIGMOD Record 34, 48–53 (2005)
6. Seglen, P.O.: The skewness of science. JASIS 43(9), 628–638 (1992)
7. Thornley, C.V., Johnson, A.C., Smeaton, A.F., Lee, H.: The scholarly impact of TRECVid (2003–2009). JASIST 62(4), 613–627 (2011)
8. Tsikrika, T., Seco de Herrera, A.G., Müller, H.: Assessing the scholarly impact of ImageCLEF. In: Forner, P., Gonzalo, J., Kekäläinen, J., Lalmas, M., de Rijke, M. (eds.) CLEF 2011. LNCS, vol. 6941, pp. 95–106. Springer, Heidelberg (2011)
9. Voorhees, E.M.: The philosophy of information retrieval evaluation. In: Peters, C., Braschler, M., Gonzalo, J., Kluck, M. (eds.) CLEF 2001. LNCS, vol. 2406, pp. 355–370. Springer, Heidelberg (2002)