

ImageCLEF 2020: Multimedia Retrieval in Lifelogging, Medical, Nature, and Internet Applications

Bogdan Ionescu¹, Henning Müller², Renaud Péteri³, Duc-Tien Dang-Nguyen⁴,
Liting Zhou⁵, Luca Piras⁶, Michael Riegler⁷, Pål Halvorsen⁷, Minh-Triet Tran⁸,
Mathias Lux⁹, Cathal Gurrin⁵, Jon Chamberlain¹⁰, Adrian Clark¹⁰, Antonio
Campello¹¹, Alba G. Seco de Herrera¹⁰, Asma Ben Abacha¹², Vivek Datla¹³,
Sadid A. Hasan¹⁴, Joey Liu¹³, Dina Demner-Fushman¹², Obioma Pelka¹⁵,
Christoph M. Friedrich¹⁵, Yashin Dicente Cid¹⁶, Serge Kozlovski¹⁷, Vitali
Liauchuk¹⁷, Vassili Kovalev¹⁷, Raul Berari¹⁸, Paul Brie¹⁸, Dimitri Fichou¹⁸,
Mihai Dogariu¹, Liviu Daniel Stefan¹, and Mihai Gabriel Constantin¹

¹ University Politehnica of Bucharest, Romania bogdan.ionescu@upb.ro

² University of Applied Sciences Western Switzerland (HES-SO), Switzerland

³ University of La Rochelle, France

⁴ University of Bergen, Norway

⁵ Dublin City University, Ireland

⁶ Pluribus One & University of Cagliari, Italy

⁷ SimulaMet, Norway

⁸ University of Science, Vietnam

⁹ Klagenfurt University, Austria

¹⁰ University of Essex, UK

¹¹ Wellcome Trust, UK

¹² National Library of Medicine, USA

¹³ Philips Research Cambridge, USA

¹⁴ CVS Health, USA

¹⁵ University of Applied Sciences and Arts Dortmund, Germany

¹⁶ University of Warwick, UK

¹⁷ United Institute of Informatics Problems, Belarus

¹⁸ teleportHQ, Romania

Abstract. This paper presents an overview of the 2020 ImageCLEF lab that will be organized as part of the Conference and Labs of the Evaluation Forum — CLEF Labs 2020 in Thessaloniki, Greece. ImageCLEF is an ongoing evaluation initiative (run since 2003) that promotes the evaluation of technologies for annotation, indexing and retrieval of visual data with the aim of providing information access to large collections of images in various usage scenarios and domains. In 2020, the 18th edition of ImageCLEF will organize four main tasks: (i) a *Lifelog* task (videos, images and other sources) about daily activity understanding, retrieval and summarization, (ii) a *Medical* task that groups three previous tasks (caption analysis, tuberculosis prediction, and medical visual question answering) with new data and adapted tasks, (iii) a *Coral* task about segmenting and labeling collections of coral images for 3D modeling, and

a new (iv) *Web user interface* task addressing the problems of detecting and recognizing hand drawn website UIs (User Interfaces) for generating automatic code. The strong participation, with over 235 research groups registering and 63 submitting over 359 runs for the tasks in 2019 shows an important interest in this benchmarking campaign. We expect the new tasks to attract at least as many researchers for 2020.

Keywords: lifelogging retrieval and summarization · medical image classification · coral image segmentation and classification · recognition of hand drawn website UIs · ImageCLEF benchmarking · annotated data

1 Introduction

The ImageCLEF evaluation campaign was started as part of the CLEF (Cross Language Evaluation Forum) in 2003 [7, 8]. It has been held every year since then and delivered many results in the analysis and retrieval of images [21, 20]. Medical tasks started in 2004 and have in some years been the majority of the tasks in ImageCLEF [19, 18].

The objectives of ImageCLEF have always been the multilingual or language-independent analysis of visual content. A focus has often been on multimodal data sets, so combining images with structured information, free text or other information that helps in the decision making.

Since 2018, ImageCLEF uses the crowdAI (now migrated to Aicrowd¹⁹ starting with 2020) platform to distribute the data and receive the submitted results. The system allows having an online leader board and gives the possibility to keep data sets accessible beyond competition, including a continuous submission of runs and addition to the leader board.

Over the years, ImageCLEF and also CLEF have shown a strong scholarly impact that was captured in [27, 28]. This underlines the importance of evaluation campaigns for disseminating best scientific practices.

In the following, we introduce the four tasks that are going to run in the 2020 edition²⁰, namely: ImageCLEFlifelog, ImageCLEFmedical, ImageCLEFcoral, and the new ImageCLEFdrawnUI. Figure 1 captures with a few images the specificity of the tasks.

2 ImageCLEFlifelog

The main goal of the Lifelog task since its first edition [9] has been to advance the state-of-the-art research in lifelogging as an application of information retrieval. Different personal devices, such as smartphones, video cameras as well as wearable devices that allow collection of different data about our daily life are available. Large amounts of data are created by these devices containing videos,

¹⁹ <https://www.aicrowd.com/>

²⁰ <https://www.imageclef.org/2020>

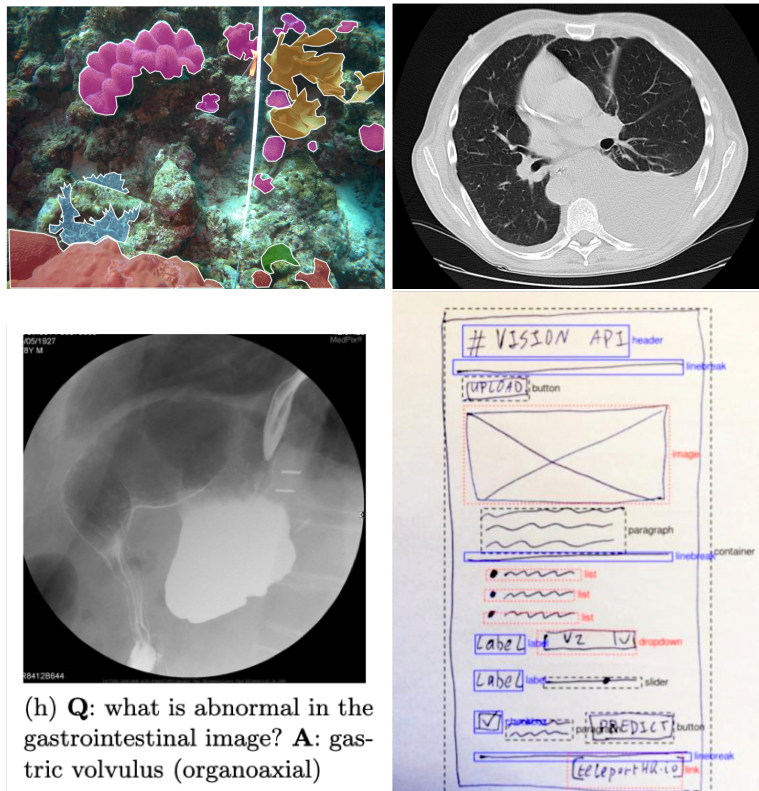


Fig. 1. Sample images from (left to right, top to bottom): ImageCLEFcoral with coral image segmentation and labeling, ImageCLEFmedical, e.g., tuberculosis prediction and Visual Question Answering, and ImageCLEFdrawnUI with recognition of hand drawn website UIs.

images, audio and sensor data. To be able to organize such a vast amount of data, there is a clear need of systems that can do this automatically.

As in the previous three editions, the task focuses mainly on images. The 2020 task will again be split into two subtasks: the lifelog moment retrieval and a new sports performance lifelog task. The first subtask includes new and enriched data, focusing on daily living activities and the chronological order of the moments. The second subtask provides a completely new dataset for assessing sports performance.

For the *Lifelog Core Task: Lifelog Moment Retrieval* the participants are required to retrieve several predefined activities in a lifelogger's life. For example, they are asked to return the relevant moments for the query "Find the moment(s) when the lifelogger was having a beer on the beach with his/her friends". Particular attention will be paid to the diversification of the selected moments with respect to the target scenario. To make the task possible and

interesting a rich multimodal dataset will be used. The data are completely new and contain about 4.5 months of data from three lifeloggers, 1,500-2,500 images per day, visual concepts, semantic content, biometric information, music listening history and computer usage.

The other task, *Lifelog Task: Sports Performance Lifelog* (SPLL, 1st edition), is completely new in terms of data and topic. Teams are required to predict the expected performance (e.g., estimated finishing time, average heart rate and other performance measurements) for a non-professional athlete who trained for a sports event. For the task, a new dataset is provided containing information collected from 20-24 people that train for a 5 km run. Objective sensor data is collected using the FitBit Versa 2 sport watch²¹; subjective wellness, training load and injury data is collected using the PMSYS system²²; and information about meals, drinks, medication, etc. is collected using Google Forms. The data contain information about daily sleeping patterns, daily heart rate, sport activities, logs of food consumed during the training period (from at least 2 participants) and self reported data like mood, stress, fatigue, readiness to train and other measurements also used for professional soccer teams. The data are collected over a period of four to five month. The copyright and ethical approval to release the data are obtained by the task organizers. For the sports task data, we have the data approved by the Norwegian Center for Research Data²³. For assessing the performance of the approaches, classic metrics will be used, e.g., precision, cluster recall (to account for the diversification), etc. For this sports task, we will also utilize metrics such as Mean Absolute Error (MAE) and Root mean squared error (RMSE).

3 ImageCLEFmedical

The ImageCLEF medical task has been running every year since 2004 [22]. In 2020, it will follow a similar format as in the previous edition [18] containing the same three subtasks with some modifications. The three tasks will be: *tuberculosis analysis* [10, 12, 11], *figure caption analysis* [13, 17, 23], and *Visual Question Answering* [16, 2].

The *tuberculosis task* will use, as in previous editions, Computed Tomography (CT) scans of patients with tuberculosis and more clinical data. In this edition, the task will concentrate only on generating an automatic report based on the CT and not assessing a TB severity score. The new report will be more detailed than in the previous edition, containing more specific information, such as in which region each TB-related finding is located.

The *caption analysis task* will include more data compared to 2019. In 2020, an extension of the Radiology Objects in Context (ROCO) [24] data set is used and manually curated to reduce the data variability. The collection includes images from the medical literature including caption information, concepts and

²¹ <https://www.fitbit.com/versa>

²² <https://forzasys.com/pmsys.html>

²³ <https://nsd.no>

7 sub-classes denoting the image radiology modality. The task concentrates on extracting Unified Medical Language System (UMLS®) Concept Unique Identifiers (CUIs) and can also be used as a first step towards the *Medical Visual Question Answering (VQA-Med)* task.

The *medical Visual Question Answering (VQA-Med) task* poses a challenging problem that involves both natural language processing and computer vision. In continuation of the two previous editions, the task consists of answering a natural language question from the visual content of an associated radiology image. VQA-Med 2020 will focus further on questions about abnormalities and will include a new subtask on visual question generation from radiology images.

4 ImageCLEFcoral

Coral reefs are important ecosystems because they are the most biodiverse parts of the oceans. However, corals thrive in narrow temperature ranges and ocean warming trends, among other factors, indicate that many of them will be lost within the next 30 years [3]. This would be a catastrophe, not only because of the extinction of many of the marine species they host but also because they provide an income and an essential food source to the people who live nearby [4, 25]. Monitoring changes in reef composition and structural complexity on a large-scale is crucial to understanding and prioritizing conservation efforts.

Key to conservation work is knowledge of the state of reefs. Autonomous underwater vehicles are able to collect large amounts of data, more than can be annotated by a human. Although there have been promising attempts at automatically annotating imagery of reefs for complexity and benthic composition [26, 15], it is fair to say that the problem is far from being solved. The aim of this competition is to encourage researchers to improve techniques for automatically identifying areas of interest and label them in a way that helps marine biologists and ecologists.

Following the success of the first edition of the ImageCLEFcoral task [5], in 2020, participants are required to devise and implement algorithms for automatically annotating regions in a collection of images with types of benthic substrate, such as hard coral or sponge. The dataset comprises 440 human-annotated training images and a further 200 unseen test images of a region of coral reef in Indonesia. The images were captured in high-quality JPEG format using an innovative underwater image capture system developed at the Marine Technology Research Unit at the University of Essex, UK.

The ground truth annotations of the training and test sets were made largely by marine biology MSc students at Essex and checked by an experienced coral reef researcher. The annotations were performed using a web-based tool developed in a collaborative project with London-based company Filament Ltd which allowed many people to work concurrently and which was carefully designed to be simple and quick to use; this proved so effective that we are exploring whether the tool can be made publicly available for other tasks in the future.

As in the first edition, algorithmic performance will be evaluated on the unseen test data using the intersection over union metric popularized in the PASCAL VOC²⁴ exercise. This computes the area of intersection of the output of an algorithm and the corresponding ground truth, normalizing that by the area of their union to ensure its maximum value remains bounded.

5 ImageCLEFdrawnUI

The user interface (UI) is the space where interactions between humans and computers occur. The increasing dependence of web and mobile applications have led many enterprises to increase the priority of developing user interfaces, in an effort to improve the overall user experience. Currently, the performance of any modern digital product is strongly correlated to the quality and usability of its user interface. However, building a user interface for digital applications is a complex process involving the interaction between multiple specialists, each with its own specific domain knowledge.

Generally, a business owner is setting a business goal. Then, a project owner or product owner is refining the requirements and builds a prototype of the application using wireframes. Once validated, the designer transforms the wireframes into designs which are transformed into code by the developer. This process is time consuming and expensive. Moreover, as more people get involved, the process is increasingly error prone. In addition, user interface experts are in limited supply. Globally, there are about 22 million developers²⁵, among which only 10 million are estimated to be also JavaScript UI developers²⁶.

Recently the use of machine learning to facilitate this process has been demonstrated as a viable solution. In 2018, pix2code, a machine learning based approach to generate low fidelity domain specific languages from screenshots, was published and open sourced [1]. Also, in 2018 Chen Chunyang et al. created their own dataset from android apps with 185,277 pairs of UI images and GUI skeletons. The dataset and code were also open-sourced [6].

In this context, in the 2020 ImageCLEFdrawnUI task, given a set of images of hand drawn UIs, participants are required to develop machine learning techniques that are able to predict the exact position and type of UI elements. The provided dataset consists of 3,000 hand drawn images inspired from mobile application screenshots and actual web pages containing 1,000 different templates. Each image was manually labelled with the positions of the bounding boxes corresponding to each UI element and its type. To avoid any ambiguity, a pre-defined shape dictionary with 21 classes is used (e.g., paragraph, label, header). The performance of the algorithms will be evaluated using the standard mean Average Precision over IoU .5, commonly used in object detection [14].

²⁴ <http://host.robots.ox.ac.uk/pascal/VOC/>

²⁵ <http://evansdata.com>

²⁶ <http://appdeveloper magazine.com>

6 Conclusions

In this paper, we present an overview of the upcoming ImageCLEF 2020 campaign. ImageCLEF has organized many tasks in a variety of domains over the past 18 years, from general stock photography, medical and biodiversity data to multimodal lifelogging. The focus has always been on language independent or multi-lingual approaches and most often on multimodal data analysis. 2020 has a set of interesting tasks that are expected to again draw a large number of participants. As in 2019, the focus for 2020 has been on the diversity of applications and on creating clean data sets to provide a solid basis for the evaluations.

Acknowledgement

Mihai Dogariu and Liviu Daniel Stefan work has been funded by the Operational Programme Human Capital of the Ministry of European Funds through the Financial Agreement 51675/09.07.2019, SMIS code 125125.

References

1. Beltramelli, T.: pix2code : Generating Code from a Graphical User Interface Screenshot. Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems pp. 1–9 (2018)
2. Ben Abacha, A., Hasan, S.A., Datla, V.V., Liu, J., Demner-Fushman, D., Müller, H.: Vqa-med: Overview of the medical visual question answering task at imageclef 2019. In: Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, Lugano, Switzerland, September 9-12, 2019. (2019), http://ceur-ws.org/Vol-2380/paper_272.pdf
3. Birkeland, C.: Global status of coral reefs: In combination, disturbances and stressors become ratchets. In: World Seas: an Environmental Evaluation, pp. 35–56. Elsevier (2019)
4. Brander, L.M., Rehdanz, K., Tol, R.S., Van Beukering, P.J.: The economic impact of ocean acidification on coral reefs. *Climate Change Economics* **3**(01), 1250002 (2012)
5. Chamberlain, J., Campello, A., Wright, J.P., Clift, L.G., Clark, A., García Seco de Herrera, A.: Overview of ImageCLEFcoral 2019 task. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, vol. 2380. CEUR-WS.org (2019)
6. Chen, C., Su, T., Meng, G., Xing, Z., Liu, Y.: From UI Design Image to GUI Skeleton : A Neural Machine Translator to Bootstrap Mobile GUI Implementation. *International Conference on Software Engineering* **6** (2018)
7. Clough, P., Müller, H., Sanderson, M.: The CLEF 2004 cross-language image retrieval track. In: Peters, C., Clough, P., Gonzalo, J., Jones, G.J.F., Kluck, M., Magnini, B. (eds.) *Multilingual Information Access for Text, Speech and Images: Result of the fifth CLEF evaluation campaign*. Lecture Notes in Computer Science (LNCS), vol. 3491, pp. 597–613. Springer, Bath, UK (2005)
8. Clough, P., Sanderson, M.: The CLEF 2003 cross language image retrieval task. In: *Proceedings of the Cross Language Evaluation Forum (CLEF 2003)* (2004)

9. Dang-Nguyen, D.T., Piras, L., Riegler, M., Boato, G., Zhou, L., Gurrin, C.: Overview of ImageCLEFlifelog 2017: Lifelog Retrieval and Summarization. In: Cappellato, L., Ferro, N., Goeuriot, L., Mandl, T. (eds.) CLEF 2017 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1866/> (2017)
10. Dicente Cid, Y., Kalinovsky, A., Liauchuk, V., Kovalev, V., Müller, H.: Overview of ImageCLEFtuberculosis 2017 - predicting tuberculosis type and drug resistances. In: CLEF2017 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <<http://ceur-ws.org>>, Dublin, Ireland (September 11-14 2017)
11. Dicente Cid, Y., Liauchuk, V., Klimuk, D., Tarasau, A., Kovalev, V., Müller, H.: Overview of ImageCLEFtuberculosis 2019 - automatic ct-based report generation and tuberculosis severity assessment. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <<http://ceur-ws.org>>, Lugano, Switzerland (September 9-12 2019)
12. Dicente Cid, Y., Liauchuk, V., Kovalev, V., Müller, H.: Overview of ImageCLEFtuberculosis 2018 - detecting multi-drug resistance, classifying tuberculosis type, and assessing severity score. In: CLEF2018 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <<http://ceur-ws.org>>, Avignon, France (September 10-14 2018)
13. Eickhoff, C., Schwall, I., García Seco de Herrera, A., Müller, H.: Overview of ImageCLEFcaption 2017 - the image caption prediction and concept extraction tasks to understand biomedical images. In: CLEF2017 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <<http://ceur-ws.org>>, Dublin, Ireland (September 11-14 2017)
14. Everingham, M., Gool, L.V., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes (VOC) Challenge. *Int J Comput Vis* **88**, 303–338 (2010). <https://doi.org/10.1007/s11263-009-0275-4>
15. Gonzalez-Rivero, M., Bongaerts, P., Beijbom, O., Pizarro, O., Friedman, A., Rodriguez-Ramirez, A., Upcroft, B., Laffoley, D., Kline, D., Bailhache, C., Vevers, R., Hoegh-Guldberg, O.: The catlin seaview survey—kilometre-scale seascape assessment, and monitoring of coral reef ecosystems. *Aquatic Conservation: Marine and Freshwater Ecosystems* **24**, 184–198 (11 2014)
16. Hasan, S.A., Ling, Y., Farri, O., Liu, J., Lungren, M., Müller, H.: Overview of the ImageCLEF 2018 medical domain visual question answering task. In: CLEF2018 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <<http://ceur-ws.org>>, Avignon, France (September 10-14 2018)
17. García Seco de Herrera, A., Eickhoff, C., Andrearczyk, V., Müller, H.: Overview of the ImageCLEF 2018 caption prediction tasks. In: CLEF2018 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <<http://ceur-ws.org>>, Avignon, France (September 10-14 2018)
18. Ionescu, B., Müller, H., Péteri, R., Cid, Y.D., Liauchuk, V., Kovalev, V., Klimuk, D., Tarasau, A., Abacha, A.B., Hasan, S.A., Datla, V., Liu, J., Demner-Fushman, D., Dang-Nguyen, D.T., Piras, L., Riegler, M., Tran, M.T., Lux, M., Gurrin, C., Pelka, O., Friedrich, C.M., de Herrera, A.G.S., Garcia, N., Kavallieratou, E., del Blanco, C.R., Rodríguez, C.C., Vasilopoulos, N., Karampidis, K., Chamberlain, J., Clark, A., Campello, A.: ImageCLEF 2019: Multimedia retrieval in medicine, lifelogging, security and nature. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the 10th International Conference of the CLEF Association (CLEF 2019), vol. 11438. LNCS Lecture Notes in Computer Science, Springer, Lugano, Switzerland (September 9-12 2019)

19. Ionescu, B., Müller, H., Villegas, M., de Herrera, A.G.S., Eickhoff, C., Andreczyk, V., Cid, Y.D., Liauchuk, V., Kovalev, V., Hasan, S.A., Ling, Y., Farri, O., Liu, J., Lungren, M., Dang-Nguyen, D.T., Piras, L., Riegler, M., Zhou, L., Lux, M., Gurrin, C.: Overview of ImageCLEF 2018: Challenges, datasets and evaluation. In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Ninth International Conference of the CLEF Association (CLEF 2018)*, LNCS Lecture Notes in Computer Science, Springer, Avignon, France (September 10-14 2018)
20. Kalpathy-Cramer, J., García Seco de Herrera, A., Demner-Fushman, D., Antani, S., Bedrick, S., Müller, H.: Evaluating performance of biomedical image retrieval systems: Overview of the medical image retrieval task at ImageCLEF 2004–2014. *Computerized Medical Imaging and Graphics* **39**(0), 55 – 61 (2015)
21. Müller, H., Clough, P., Deselaers, T., Caputo, B. (eds.): *ImageCLEF – Experimental Evaluation in Visual Information Retrieval*, The Springer International Series On Information Retrieval, vol. 32. Springer, Berlin Heidelberg (2010)
22. Müller, H., Kalpathy-Cramer, J., García Seco de Herrera, A.: Experiences from the imageclef medical retrieval and annotation tasks. In: *Information Retrieval Evaluation in a Changing World*, pp. 231–250. Springer (2019)
23. Pelka, O., Friedrich, C.M., García Seco de Herrera, A., Müller, H.: Overview of the ImageCLEFmed 2019 concept prediction task. In: *CLEF2019 Working Notes. CEUR Workshop Proceedings*, vol. 2380. CEUR-WS.org <<http://ceur-ws.org>>, Lugano, Switzerland (September 09-12 2019)
24. Pelka, O., Koitka, S., Rückert, J., Nensa, F., Friedrich, C.M.: Radiology Objects in COntext (ROCO): a multimodal image dataset. In: *Proceedings of the Third International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis (LABELS 2018)*, Held in Conjunction with MICCAI 2018. vol. 11043, pp. 180–189. LNCS Lecture Notes in Computer Science, Springer, Granada, Spain (September 16 2018)
25. Speers, A.E., Besedin, E.Y., Palardy, J.E., Moore, C.: Impacts of climate change and ocean acidification on coral reef fisheries: An integrated ecological–economic model. *Ecological economics* **128**, 33–43 (2016)
26. Stokes, M., Deane, G.: Automated processing of coral reef benthic images. *Limnology and Oceanography Methods* **7**, 157–168 (2009)
27. Tsirikika, T., García Seco de Herrera, A., Müller, H.: Assessing the scholarly impact of ImageCLEF. In: *CLEF 2011*. pp. 95–106. Springer Lecture Notes in Computer Science (LNCS) (sep 2011)
28. Tsirikika, T., Larsen, B., Müller, H., Endrullis, S., Rahm, E.: The scholarly impact of CLEF (2000–2009). In: *Information Access Evaluation. Multilinguality, Multimodality, and Visualization*, pp. 1–12. Springer (2013)