



Anna Corazza, Simonetta Montemagni and Giovanni Semeraro (dir.)

Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it 2016 5-6 December 2016, Napoli

Accademia University Press

Gamification for IR: The Query Aspects Game

Giorgio Maria Di Nunzio, Maria Maistro and Daniel Zilio

DOI: 10.4000/books.aaccademia.1763

Publisher: Accademia University Press

Place of publication: Torino

Year of publication: 2016

Published on OpenEdition Books: July 26, 2017

Series: Collana dell'Associazione Italiana di Linguistica Computazionale

Electronic EAN: 9788899982546



<http://books.openedition.org>

Electronic reference

DI NUNZIO, Giorgio Maria ; MAISTRO, Maria ; and ZILIO, Daniel. *Gamification for IR: The Query Aspects Game* In: *Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it 2016: 5-6 December 2016, Napoli* [online]. Torino: Accademia University Press, 2016 (generated 05 octobre 2023). Available on the Internet: <<http://books.openedition.org/aaccademia/1763>>. ISBN: 9788899982546. DOI: <https://doi.org/10.4000/books.aaccademia.1763>.



The text only may be used under licence CC BY-NC-ND 4.0. All other elements (illustrations, imported files) are "All rights reserved", unless otherwise stated.

Gamification for IR: The Query Aspects Game

Giorgio Maria Di Nunzio

Dept. of Inf. Eng. (DEI)
University of Padua, Italy
Via Gradenigo 6/a 35131

dinunzio@dei.unipd.it

Maria Maistro

Dept. of Inf. Eng. (DEI)
University of Padua, Italy
Via Gradenigo 6/a 35131

maistro@dei.unipd.it

Daniel Zilio

Dept. of Inf. Eng. (DEI)
University of Padua, Italy
Via Gradenigo 6/a 35131

daniel.zilio.3@studenti.unipd.it

Abstract

English.

The creation of a labelled dataset for Information Retrieval (IR) purposes is a costly process. For this reason, a mix of crowd-sourcing and active learning approaches have been proposed in the literature in order to assess the relevance of documents of a collection given a particular query at an affordable cost. In this paper, we present the design of the gamification of this interactive process that draws inspiration from recent works in the area of gamification for IR. In particular, we focus on three main points: i) we want to create a set of relevance judgements with the least effort by human assessors, ii) we use interactive search interfaces that use game mechanics, iii) we use Natural Language Processing (NLP) to collect different aspects of a query.¹

Italiano

La creazione di una collezione sperimentale per l'Information Retrieval (IR) è un processo costoso sia dal punto di vista economico che in termini di sforzo umano. Per ridurre i costi legati all'attribuzione dei giudizi di rilevanza ai documenti di una collezione, sono stati proposti approcci che integrano tecniche di crowd-sourcing e active learning. In questo paper viene presentata un'idea basata sull'utilizzo della gamification ('ludicizzazione') in IR per l'attribuzione di giudizi di rilevanza in maniera semi-automatica.

In particolare, ci focalizzeremo su tre aspetti principali: i) si vuole creare una collezione in modo che l'assegnazione dei giudizi da parte dei valutatori richieda il minor sforzo possibile, ii) per mezzo di un'interfaccia che utilizza dinamiche di gioco iii) insieme a tecniche di NLP per la riformulazione della query.

1 Introduction

In Information Retrieval (IR), the performance of a system is evaluated using experimental test collections. These collections consist of a set of documents, a set of queries, and a set of relevance judgements, where each judgement explains whether a document is relevant or not to each query. The creation of relevance judgements is a costly, time-consuming, and non-trivial task; for these reasons, the interest in approaches that generate relevance judgements with the least amount of effort has increased in IR and related areas (i.e., supervised Machine Learning (ML) algorithms). In the last years, mixed approaches that use crowdsourcing (Ho et al., 2013) and active learning (Settles, 2011) have shown that it is possible to create annotated datasets at affordable costs. Specifically, crowdsourcing has been part of the IR toolbox as a cheap and fast mechanism to obtain labels for system evaluation. However, successful deployment of crowdsourcing at scale involves the adjustment of many variables, a very important one being the number of assessors needed per task, as explained in (Abraham et al., 2016).

1.1 Search Diversification and Query Reformulation

The effectiveness of a search and the satisfaction of users can be enhanced through providing various results of a search query in a certain order of relevance and concern. The technique used

¹This paper is partially an extended abstract of the paper "Gamification for Machine Learning: The Classification Game" presented at the GamifIR 2016 Workshop co-located with SIGIR 2016 (Di Nunzio et al., 2016)

to avoid presenting similar, though relevant, results to the user is known as a diversification of search results (Abid et al., 2016). While existing research in search diversification offers several solutions for introducing variety into the results, the majority of such work is based on the assumption that a single relevant document will fulfil a user’s information need, making them inadequate for many informational queries. In (Welch et al., 2011), the authors propose a model to make a tradeoff between a user’s desire for multiple relevant documents, probabilistic information about an average user’s interest in the subtopics of a multifaceted query, and uncertainty in classifying documents into those subtopics.

Most information retrieval systems operate by performing a single retrieval in response to a query. Effective results sometimes require several manual reformulations by the user or semi-automatic reformulations assisted by the system. Diaz presents an approach to automatic query reformulation which combines the iterated nature of human query reformulation with the automatic behavior of pseudo relevance feedback (Diaz, 2016). In (Azzopardi, 2009), the author proposes a method for generating queries for ad-hoc topics to provide the necessary data for this comprehensive analysis of query performance. Bailey et al. explore the impact of widely differing queries that searchers construct for the same information need description. By executing those queries, we demonstrate that query formulation is critical to query effectiveness (Bailey et al., 2015).

1.2 Gamification in IR

Gamification is defined as “the use of game design elements in non-game contexts” (Deterding et al., 2011), i.e. typical game elements are used for purposes different from their normal expected employment. Nowadays, gamification spreads through a wide range of disciplines and its applications are implemented in different and various aspects of scientific fields of study. For instances, gamification is applied to learning activities (Kotini and Tzelepi, 2015; Kapp, 2012), business and enterprise (Jurado et al., 2015; Stanculescu et al., 2016; Thom et al., 2012) and medicine (Eickhoff, 2014; Chen and Pu, 2014).

IR has recently dealt with gamification, as witnessed by the Workshop on Gamification for Information Retrieval (GamifIR) in 2014, 2015 and

2016. In (Galli et al., 2014), the authors describe the fundamental elements and mechanics of a game and provide an overview of possible applications of gamification to the IR process. In (Shovman, 2014), approaches to properly gamify Web search are presented, i.e. making the search of information and the scanning of results a more enjoyable activity. Actually, many proposals of game applied to different aspects of IR have been presented. For example in (Maltzahn et al., 2014), the authors describes a game that turns document tagging into the activity of taking care of a garden, with the aim of managing private archives. A method to obtain ranking of images by utilizing human computation through a gamified web application is proposed in (Lux et al., 2014). Fort et al. introduce a strategy to gamify the annotation of a French corpora (Fort et al., 2014).

In this paper, we want to apply game mechanics to the problem of relevance assessment with three goals in mind: i) we want to create a set of relevance judgements with the least effort by human assessors, ii) we use interactive search interfaces that use game mechanics, iii) we use NLP to collect different aspects of a query. In this context, we can define our web application as a Game with a Purpose (GWAP), that is a game which presents some purposes, usually boring and dull for people, within an entertaining setting, in order to make them enjoyable and to solve problem with the aid of human computation. The design and the implementation of this interactive interface will be used as a post-hoc analysis of two Text REtrieval Conference (TREC)² 2016 tracks, namely the Total Recall Track and the Dynamic Domain Track. These two tracks are interesting for our problem since they both re-create a situation where we need to find the best set (or the total amount) of relevant documents with the minimum effort by the assessor that has to judge the documents proposed by the system given an information need.

2 Design of the Experiment

In this first pilot study, we will implement a simple game based on a visual interpretation of probabilistic classifiers (Di Nunzio, 2014; Di Nunzio, 2009; Di Nunzio and Sordoni, 2012). The game consists in separating two sets of colored points on a two-dimensional plane by means of a straight line, as shown in Figure 1. Despite its simplicity,

²<http://trec.nist.gov>

this very abstract scenario received a good feedback by kids of primary schools who tested it during the European Researcher’s Night at the University of Padua³. The next step will be to design and implement the game with real game development platforms like, for example, Unity⁴ or Marmalade⁵.

2.1 The Classification Game

The ‘original game’ (Di Nunzio et al., 2016) is based on the two-dimensional representation of probabilities (Di Nunzio, 2014; Singh and Raj, 2004), which is a very intuitive way of presenting the problem of classification on a two-dimensional space. Given two classes c_1 and c_2 , an object o is assigned to category c_1 if the following inequality holds:

$$\underbrace{P(o|c_2)}_y < m \underbrace{P(o|c_1)}_x + q \quad (1)$$

where $P(o|c_1)$ and $P(o|c_2)$ are the likelihoods of the object o given the two categories, while m and q are two parameters that depend on the misclassification costs that can be assigned by the user to compensate for either the unbalanced classes situation or different class costs.

If we interpret the two likelihoods as two coordinates x and y of a two dimensional space, the problem of classification can be studied on a two-dimensional plot. The decision of the classification is represented by the ‘line’ $y = mx + q$ that splits the plane into two parts, therefore all the points that fall ‘below’ this line are classified as objects that belong to class c_1 (see Figure 1 for an example). Without entering into the mathematical details of this approach (Di Nunzio, 2014), the basic idea of the game is that the players can adapt the two parameters m and q in order to optimize the separation of points and, at the same time, can use their resources to improve the estimate of the two likelihoods by buying training data, and/or add more points to the plot, by buying validation data.

3 The Query Aspects Game

The classification game can be easily adjusted into a relevance assessment game if the two classes are ‘relevant’ and ‘non-relevant’ (we assume only binary relevance assessment for the moment). How-

ever, while in the classification game we already have a set of labelled documents and the goal is to find the optimal classifier, in this new game we need to find the relevant documents. To this purpose, we will follow the idea of the works described in the following subsections: i) building assessment by varying the description of the information need, ii) using an interactive interface that suggests the amount of relevant information that has to be judged, iii) using NLP approaches to generate variations of a query.

3.1 Building Relevance Assessments With Query Aspects

In (Efron, 2009), the author presents a method for creating relevance judgments without explicit relevance assessments. The idea is to create different ‘aspects’ of a query: given a query q and a set of documents D , the same information need that generated q could also generate other queries that focus on another aspects of the same need. A query aspect is an articulation of a searcher’s information need which might be a re-elaboration (for example, rephrasing, specification, or generalization) of the query. By generating several queries related to an information need and running each of these against our document collection, we can create a pool of results where each result set pertains to a particular aspect of the information need with a limited human effort.

In practice, in order to build a set of relevance assessments for q , we generate a number of query aspects using a single IR system. Then, the union of the top k documents retrieved for each aspect constitutes a list of pseudo-relevance assessments for the query q .

3.2 An Interactive Interface to Generate Query Aspects

Building different aspects of the same information need is not an easy task. As explained in (Umemoto et al., 2016), searchers often cannot easily come up with effective queries for collecting documents that cover diverse aspects. In general, experts that have to search for relevant documents usually have to issue more queries to complete the tasks if search engines return few documents relevant to unexplored aspects. Moreover, quitting this tasks too early without in-depth exploration prevents searchers from finding essential information.

³<http://www.venetonight.it/padova/>

⁴<https://unity3d.com>

⁵<https://www.madewithmarmalade.com/>

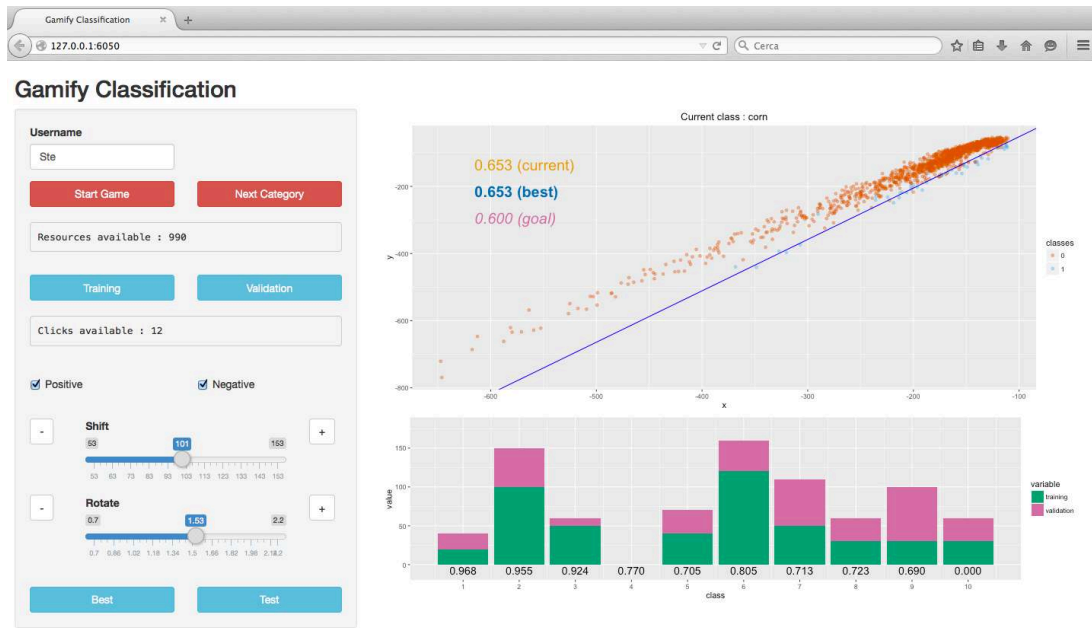


Figure 1: Layout of the original “classification game” that will be adapted to the “query aspects game”.

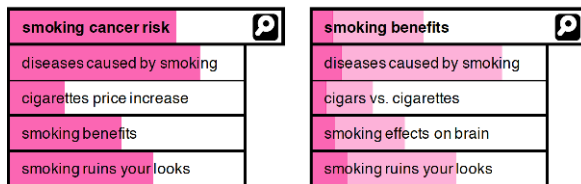


Figure 2: Scentbars and the visualization of missing information. Figure from (Umemoto et al., 2016)

Umemoto et al. propose an interactive interface, named ScentBar, that helps searchers to visualize the amount of missing information for both the search query and suggestion queries in the form of a stacked bar chart. The interface, a portion of which is shown in Figure 2, visualizes an estimate of missing information for each aspect of a query that could be obtained by the searcher. When the user collects new information during the browsing of the results, the bars of the different query aspects change color to indicate the amount of effort that the system estimates necessary to find most of the relevant information. The estimates of the required effort to complete a task are formalized as as a set-wise metric were the gain for each aspects is represented by a conditional probability.

3.3 Using NLP to Generate Query Aspects

The last part of the design of the query aspects game involves the use of natural language process-

ing techniques to propose variations of a query to express the same information need. This problem has been studied for more than twenty years in IR. In (Strzalkowski et al., 1997), the authors discuss how the simplest word-based representations of content, while relatively better understood, are usually inadequate since single words are rarely specific enough for accurate discrimination. Consequently, a better method is to identify groups of words that create meaningful phrases, especially if these phrases denote important concepts in the domain.

Some examples of advanced techniques of phrase extraction, including extended N-grams and syntactic parsing, attempt to uncover concepts, which would capture underlying semantic uniformity across various surface forms of expression. Syntactic phrases, for example, appear reasonable indicators of content since they can adequately deal with word order changes and other structural variations. In the literature, there are examples of query reformulation using NLP approaches for example to the modification and/or expansion of both parts thematic and geospatial that are usually recognized in a geographical query (Perea-Ortega et al., 2013), or to support the refinement of a vague, non-technical initial query into a more precise problem description (Roulland et al., 2007), or to predict search satisfaction (Has-san et al., 2013).

4 Conclusions and Future Works

In this work, we presented the requirements of the design of an interactive interface that uses game mechanics together with NLP techniques to generate variation of an information need in order to label a collection of documents. Starting from the successful experience of the gamification of a machine learning problem (Di Nunzio et al., 2016), we are preparing a new pilot study of the ‘query aspects game’ that will be used to generate relevant documents for two TREC tracks: the Total Recall track and the Dynamic Domain track. The results of this study will be available at the end of November 2016, and can be presented and discussed at the workshop.

References

- Adnan Abid, Naveed Hussain, Kamran Abid, Farooq Ahmad, Muhammad Shoaib Farooq, Uzma Farooq, Sher Afzal Khan, Yaser Daanial Khan, Muhammad Azhar Naeem, and Nabeel Sabir. 2016. A survey on search results diversification techniques. *Neural Computing and Applications*, 27(5):1207–1229.
- Ittai Abraham, Omar Alonso, Vasilis Kandylas, Rajesh Patel, Steven Shelford, and Aleksandrs Slivkins. 2016. How many workers to ask?: Adaptive exploration for collecting high quality labels. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR ’16, pages 473–482, New York, NY, USA. ACM.
- Leif Azzopardi. 2009. Query side evaluation: An empirical analysis of effectiveness and effort. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR ’09, pages 556–563, New York, NY, USA. ACM.
- Peter Bailey, Alistair Moffat, Falk Scholer, and Paul Thomas. 2015. User variability and ir system evaluation. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR ’15, pages 625–634, New York, NY, USA. ACM.
- Yu Chen and Pearl Pu. 2014. Healthytogether: Exploring social incentives for mobile fitness applications. In *Proceedings of the Second International Symposium of Chinese CHI*, Chinese CHI ’14, pages 25–34, New York, NY, USA. ACM.
- Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From Game Design Elements to Gamefulness: Defining “Gamification”. In *Proc. of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, MindTrek ’11, pages 9–15, New York, NY, USA. ACM.
- Giorgio Maria Di Nunzio and Alessandro Sordani. 2012. A Visual Tool for Bayesian Data Analysis: The Impact of Smoothing on Naive Bayes Text Classifiers. In *Proc. of the ACM SIGIR’12 conference on research and development in Information Retrieval*, Portland, OR, USA, August 12–16, 2012, page 1002.
- Giorgio Maria Di Nunzio, Maria Maistro, and Daniel Zilio. 2016. Gamification for machine learning: The classification game. In *Proceedings of the Third International Workshop on Gamification for Information Retrieval co-located with 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2016)*, Pisa, Italy, July 21, 2016., pages 45–52.
- Giorgio Maria Di Nunzio. 2009. Using Scatterplots to Understand and Improve Probabilistic Models for Text Categorization and Retrieval. *Int. J. Approx. Reasoning*, 50(7):945–956.
- Giorgio Maria Di Nunzio. 2014. A New Decision to Take for Cost-Sensitive Naïve Bayes Classifiers. *Information Processing & Management*, 50(5):653 – 674.
- Fernando Diaz. 2016. *Pseudo-Query Reformulation*, pages 521–532. Springer International Publishing, Cham.
- Miles Efron. 2009. Using multiple query aspects to build test collections without human relevance judgments. In *Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, ECIR ’09, pages 276–287, Berlin, Heidelberg. Springer-Verlag.
- Carsten Eickhoff. 2014. Crowd-powered experts: Helping surgeons interpret breast cancer images. In *Proceedings of the First International Workshop on Gamification for Information Retrieval*, GamifIR ’14, pages 53–56, New York, NY, USA. ACM.
- Karën Fort, Bruno Guillaume, and Hadrien Chastant. 2014. Creating zombilingo, a game with a purpose for dependency syntax annotation. In *Proceedings of the First International Workshop on Gamification for Information Retrieval*, GamifIR ’14, pages 2–6, New York, NY, USA. ACM.
- Luca Galli, Piero Fraternali, and Alessandro Bozon. 2014. On the Application of Game Mechanics in Information Retrieval. In *Proc. of the 1st Int. Workshop on Gamification for Information Retrieval*, GamifIR’14, pages 7–11, New York, NY, USA. ACM.
- Ahmed Hassan, Xiaolin Shi, Nick Craswell, and Bill Ramsey. 2013. Beyond clicks: query reformulation as a predictor of search satisfaction. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, CIKM ’13, pages 2019–2028, New York, NY, USA. ACM.

- Chien-Ju Ho, Shahin Jabbari, and Jennifer Wortman Vaughan. 2013. Adaptive task assignment for crowdsourced classification. In *Proceedings of the 30th International Conference on Machine Learning, ICML 2013*, volume 28 of *JMLR Proceedings*, pages 534–542. JMLR.org.
- Jose Luis Jurado, Alejandro Fernandez, and Cesar A. Collazos. 2015. Applying gamification in the context of knowledge management. In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business, i-KNOW '15*, pages 43:1–43:4, New York, NY, USA. ACM.
- Karl M Kapp. 2012. *The gamification of learning and instruction: game-based methods and strategies for training and education*. John Wiley & Sons.
- Isabella Kotini and Sofia Tzelepi. 2015. A Gamification-Based Framework for Developing Learning Activities of Computational Thinking. In Torsten Reiners and C. Lincoln Wood, editors, *Gamification in Education and Business*, pages 219–252. Springer Int. Publ., Cham.
- Mathias Lux, Mario Guggenberger, and Michael Riegler. 2014. Picturesort: Gamification of image ranking. In *Proceedings of the First International Workshop on Gamification for Information Retrieval, GamifIR '14*, pages 57–60, New York, NY, USA. ACM.
- Carlos Maltzahn, Arnav Jhala, Michael Mateas, and Jim Whitehead. 2014. Gamification of private digital data archive management. In *Proceedings of the First International Workshop on Gamification for Information Retrieval, GamifIR '14*, pages 33–37, New York, NY, USA. ACM.
- José M. Perea-Ortega, Miguel A. García-Cumbreras, and L. Alfonso Ureña López. 2013. Applying nlp techniques for query reformulation to information retrieval with geographical references. In *Proceedings of the 2012 Pacific-Asia Conference on Emerging Trends in Knowledge Discovery and Data Mining, PAKDD'12*, pages 57–69, Berlin, Heidelberg. Springer-Verlag.
- Frédéric Roulland, Aaron Kaplan, Stefania Castellani, Claude Roux, Antonietta Grasso, Karin Pettersson, and Jacki O'Neill, 2007. *Query Reformulation and Refinement Using NLP-Based Sentence Clustering*, pages 210–221. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Burr Settles. 2011. Closing the loop: Fast, interactive semi-supervised annotation with queries on features and instances. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011, 27-31 July 2011, John McIntyre Conference Centre, Edinburgh, UK, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1467–1478.
- Mark Shovman. 2014. The Game of Search: What is the Fun in That? In *Proc. of the 1st Int. Workshop on Gamification for Information Retrieval, GamifIR'14*, pages 46–48, New York, NY, USA. ACM.
- Rita Singh and Bhiksha Raj. 2004. Classification in likelihood spaces. *Technometrics*, 46(3):318–329.
- Laurentiu Catalin Stanculescu, Alessandro Bozzon, Robert-Jan Sips, and Geert-Jan Houben. 2016. Work and play: An experiment in enterprise gamification. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, CSCW '16*, pages 346–358, New York, NY, USA. ACM.
- Tomek Strzalkowski, Fang Lin, Jose Perez-Carballo, and Jin Wang. 1997. Building effective queries in natural language information retrieval. In *Proceedings of the Fifth Conference on Applied Natural Language Processing, ANLC '97*, pages 299–306, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Jennifer Thom, David Millen, and Joan DiMicco. 2012. Removing gamification from an enterprise sns. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, CSCW '12*, pages 1067–1070, New York, NY, USA. ACM.
- Kazutoshi Umemoto, Takehiro Yamamoto, and Katsumi Tanaka. 2016. Scentbar: A query suggestion interface visualizing the amount of missed relevant information for intrinsically diverse search. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '16*, pages 405–414, New York, NY, USA. ACM.
- Michael J. Welch, Junghoo Cho, and Christopher Olston. 2011. Search result diversity for informational queries. In *Proceedings of the 20th International Conference on World Wide Web, WWW '11*, pages 237–246, New York, NY, USA. ACM.