78% of women in tech feel they have to work harder than their coworkers to prove their worth.

ADVICE

Managing Research the Wiki Way

Long-term research requires a flexible and systematic approach for managing documentation about surveyed papers or carried experiments. Wikis have been used effectively in learning environments [1]. If you are interested in how to do so presented here is a methodology for managing research the wiki way, which provides an accessible framework for the organization of literature, datasets, experiments, and results. We prepared a DokuWiki instance with a series of templates, spread over a predefined namespace structure built to support and facilitate research management. We also provide a Docker-based installation¹ of a preconfigured wiki, with the required templates and extensions, along with example pages, as well as a Jupyter notebook² to analyze the work documented using the wiki.

SYSTEMATIC DOCUMENTATION

Using DokuWiki, a PHP, file-based wiki, we documented our research over four years. DokuWiki is easy to deploy and backup, while retaining historical changes to pages and supporting extensibility. We systematically organized information about literature, collections, and experiments, providing templates for reading sheets, collection descriptions, and experiment note taking and a results archive. The wiki page structure is summarized in Figure 1. We thoroughly exploited links and backlinks, establishing relations between authors or conferences, their publications and datasets (collections), and their mentions on other pages. We also linked subsets to their original datasets or re-

lated experiments among themselves. Our documentation methodology consists of the following:

1. a literature review methodology supported on the creation of a reading sheet;
2. a lightweight approach for documenting collections based on a description sheet template, which includes information about subsets and evaluation results taken from the literature;
3. and a note-taking and archival strategy for experiments and results.

LITERATURE REVIEW

We relied on an exploratory literature review approach, focusing and refining along the way, as concepts became clearer. We used academic search engines to issue queries in an attempt to solve our information needs about our research topic. The resulting publications were selected based on the title, the abstract, the conclusions, and sometimes a part of the introduction, in this order. Selected publications were then added to the wiki, along with a list of specific goals in the form of tasks, to be reviewed in order of priority regarding ongoing research work, or based on the overall relevance and informational value to the research topic. A reading sheet was

created from a template for each publication, either assigning a link based on the title, thus improving link readability, or based on the format <First Author Last Name><Year>, as an alternative to improve the ability to quickly reach surveyed publications. The reading sheet contained a standardized table of information with the fields: 1. authors, 2. journal/conference/etc., and 3. year; each linking to a page of backlinks (e.g., `phd:bibliography:author:w-bruce-croft`). It also contained reading notes organized according to the structure of the sections of the publication. We frequently included block quotes highlighting important information and a summary paragraph of the work.

COLLECTIONS

We created a page in the collections section of the wiki for each dataset that we used, contributed, or otherwise explored. To ensure consistency in the description, we prepared a template containing a table of metadata about each collection, along with a longer textual description. The main fields were: 1. source, 2. paper, 3. date, and 4. size. When appropriate, we provided collection statistics. We also included evaluation results found within the literature, relevant to us, which relied on the described dataset. This information is usually available in overview papers from the evaluation forums that provide the test collection.

EXPERIMENTS

We established a template for the documentation of experiments in the wiki, similar to the literature review and collection description. Every experiment page contained a metadata table with the following fields:

1. ID (e.g., “Experiment 1”)

We systematically organized information about literature, collections, and experiments...

¹ <https://github.com/jldevezas/research-wiki/>

² https://github.com/jldevezas/research-wiki/blob/main/notebook/research_wiki_analysis.ipynb

Figure 1. Page structure for the research management wiki.

2. Start date (e.g., “2017-10-24 16:38”)
3. End date (e.g., “ongoing”)
4. Why do it? (The motivation usual-ly in the sequence of a previous experi-ment)
5. Main strengths (expected im-provements)
6. Main weaknesses (predicted is-sues)
7. Test collection (a link to a wiki collection page)

Every experiment included a to-do list, a description table of the explored model versions, and an evaluation section with performance metrics for every version. Additionally, we added tables to describe predicted or identified chal-lenges, and dependencies that might block the experiment execution. Over time, the challenges and dependen-cies sections were deprecated in favor of the to-do list. On the other side, the archive of results from evaluation tasks was a useful practice that enabled us, for instance, to verify the calculation of effectiveness metrics. We created a sec-tion for research logs, where we added links to wiki pages under the hierarchy

of the current experiment. Every re-search log entry represented a reflec-tion on parts of the studied models, sometimes branching into follow-up or sub-experiments.

At the end of four years of work, we can safely say relying on a wiki to sys-tematically document our research work was the best decision. We hope the wiki way also helps you draft your research ideas and manage the process.

ACKNOWLEDGMENTS

We thank Professor João Correia Lopes, who is also an enthusiastic wiki user, for sharing his favorite DokuWiki plugins, many of which we adopted for the research management wiki we present here.

FUNDING

José Devezas is supported by research grant PD/BD/128160/2016, provided by the Portuguese national funding agency for science, research, and technology, Fundação para a Ciência e a Tecnologia (FCT), within the scope of Operational Program Human Capital (POCH), sup-

ported by the European Social Fund and by national funds from MCTES.

References

- [1] Ruth, A. and Houghton, L. The wiki way of learning. *Australasian Journal of Educational Technology* 25, 2 (2009).

Biographies

José Devezas has recently attained his Ph.D. in computer science from MAP-I, the doctoral program in computer science of the Universities of Minho, Aveiro, and Porto. His thesis, entitled “Graph-Based Entity-Oriented Search,” was born from his recurrent fascination with connecting data and building general models to help people solve their information needs. He has done work in several domains, including information retrieval, network science, music recommender systems, and data visualization. He has dedicated a lot of his time to the exploration of hypergraphs as a joint representation model for corpora and knowledge bases, along with the implementation of a universal ranking function, integrated into one of the first unified frameworks for information retrieval. More information can be found at <http://josedezas.com/>.

Sérgio Nunes is an assistant professor at the Department of Informatics Engineering at the Faculty of Engineering of the University of Porto (FEUP), and a senior researcher at the Centre for Information Systems and Computer Graphics at INESC TEC. He holds a Ph.D. in information retrieval [2010] focused on using temporal features for relevance estimation, and an M.Sc. in information management. His research interests are in the areas of information retrieval and web information systems, in particular in the use of temporal features for ranking, the study of information dynamics on the web, and computational journalism. More information and selected publications can be found at <https://web.fe.up.pt/~ssn/>.

DOI: 10.1145/3453117 Copyright held by author.