Relevation!: An Open Source System for Information Retrieval Relevance Assessment

Bevan Koopman
Australian e-Health Research Centre, CSIRO
Brisbane, Australia
bevan.koopman@csiro.au

Guido Zuccon
Queensland University of Technology
Brisbane, Australia
g.zuccon@qut.edu.au

ABSTRACT

Relevation! is a system for performing relevance judgements for information retrieval evaluation. Relevation! is web-based, fully configurable and expandable; it allows researchers to effectively collect assessments and additional qualitative data. The system is easily deployed allowing assessors to smoothly perform their relevance judging tasks, even remotely. Relevation! is available as an open source project at: http://ielab.github.io/relevation.

Categories and Subject Descriptors: H.3.3 [Information Storage and Retrieval]

General Terms: Measurement, Experimentation.

1. INTRODUCTION

An integral part of information retrieval (IR) evaluation is the use of standard test collections. Relevance assessments are critical to the quality of test collections and obtaining assessments is often a long and laborious task, which, in many cases, involves a large number of documents to be judged by multiple assessors. Performing relevance assessments is often a one-off task for the creators of test collections and is usually under tight time and budget constraints. As a result, collection creators often have little resources to invest in a high quality tool to aid judges in performing relevance assessments, even though such a tool could greatly aid this process. With this in mind, we have developed Relevation!, an adaptable, open source tool for performing relevance judgements.

Relevation! allows users to upload documents and queries that are then browsed by judges through a web-based interface; for each document, judges can assign a relevance label. The system also allows judges to provide qualitative feedback about the judgement process, both at query level and at document level. Relevation! is open source and uses the Model-View-Controller design pattern, making it customisable to specific user requirements. The system is written in Python using the Django Web framework, making it easily deployed and remotely available. Relevation! has already

been used to collect additional relevance assessments for the TREC Medical Records Track [1] and for the CLEF eHealth (2013 & 2014) evaluation campaigns [2].

Early work by Hawking et al. [3] developed a tool for collecting relevance assessment. However, this tool is outdated, does not provide web-based deployments and is no longer available. RAT is a more recent web-based system for obtaining relevance assessments [4] but is specific to web search engines and assumes documents are screen scraped from the web. In addition, RAT is not open source so cannot be extended or adapted to specific needs.

2. FEATURES OF Relevation!

The architecture of Relevation! is shown in Fig 1. The system comprises four main modules: (i) the setup module, (ii) the queries module, (iii) the documents module, and (iv) the judgements module.

Setup Module. This module allows users to upload their queries and the document judging pool to Relevation!. Two files are required: 1) Query file in tab separated format `QueryId [tab] QueryText`, and 2) Document judging pool using the standard TREC results file format. Note that the standard setup module can be extended with additional parsers for query files and document judging pool; for example, the deployment of Relevation! for the CLEF eHealth 2013 was extended by implementing a query file parser which accepted additional query fields like "Description", "Narrative" and "Profile".

Queries Module. Judges are presented with a list of queries currently in the system (screenshot of Fig 2(a)). For each query, the number of documents assigned to that query is displayed, along with the number of unjudged documents, giving an indication of the overall judging progress. In the screenshot, each query has a `QueryId` column which identi-
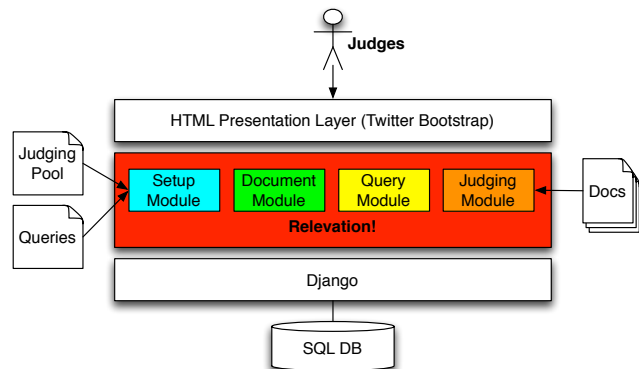

Figure 1: Architecture of Relevation!

Queries

QueryId	Text	Number of documents	Number unjudged
12	clots in jugular and HIV	114	99
17	urinary tract infection and ciprofloxacin	93	83
28	bruises and raccoon eyes and blood disease	95	95
31	endocarditis	87	77
47	sacral ulcer and care	92	92

(a) The list of queries currently in the system (query module).

12 - clots in jugular and HIV   Back to Queries

Query	Number	Document#	Status
12	1	aids.0922_12_001923	Not relevant
12	2	aids.0922_12_002121	On topic, but unreliable
12	3	aids.0922_12_002174	Highly relevant
12	4	aids.0922_12_000051	On topic, but unreliable
12	5	merck0849_12_001119	Highly relevant

(b) List of docs. assigned to single query (documents module).

(c) Assessing a document (judgements module).

Figure 2: Screenshots from the CLEF eHealth 2013 deployment of Relevation!.

fies it and an associated Text column, which is the actual keywords for that query.

Documents Module. Clicking on the Text field entry takes the judge to the next page: the list of documents associated with that query (Fig 2(b)). For each document, the assessment Status column shows the relevance label assigned to the documents (e.g., *highly relevant*, *somewhat relevant*, etc.). The status field provides the judge with a quick and easy overview of the progress of this query, as well as the collection creators with an overview of the relevance assessments for that query (e.g., distribution of relevant/irrelevant documents). The Documents# field is the filename of the particular document and is a link to the judgements page.

Once all the documents for a query are judged, the assessors can optionally provide some qualitative feedback on the particular query. A short questionnaire is presented at the bottom of the list of documents. The questionnaire can be removed or customised to suit the particular needs of the relevance assessment task.

Judgements Module. This is where judges can read a document and enter their assessment (Fig 2(c)). The top of the page gives the query keywords and description. The document contents are displayed on the lefthand side.[1] On the right hand side panel is a choice for the relevance assessment ("Not Judged" is the default judgement). Judges can also select portions of the document content and the selected text will automatically be added to the *Supporting Evidence* field. This information can be used for qualitative analysis after judging is complete or for passage based retrieval evaluation. Once assessment is complete, judges have to press the Save&Next button (not shown in screenshot), which saves the assessment and loads the next document for judging. The judgement module is customisable and different deployments of Relevation! may implement different judgements modules. For example the judgement module for CLEF eHealth 2013 also displayed narrative and profiles for the queries and used a different relevance assessment scale.

Other functionality. A script is provided for exporting relevance judgments in the standard TREC qrel format from the SQL database of Relevation! In addition, the SQL database can easily be queried to export other data, e.g., the qualitative questionnaires provided by the judges.

3. CONCLUSIONS AND FUTURE WORK

In its current version, Relevation! provides an open-source, modular, customisable system for collecting relevance assessments. In future versions of Relevation!, we plan to integrate the qrel creation tools and database querying methods within the core modules. The Setup module will be extended to allow users to configure custom relevance grades and qualitative questionnaires. Additional document parsers for the main TREC collections will also be added to this module. The Judging module will be extend to allow configurable placement of documents within the judgement interface, e.g., to support the visualisation of two documents at the time for preference judgements. Finally, we plan to incorporate a new Crowdsourcing module that allows Relevation! to outsource relevance assessment collection to workforce platforms such as Amazon Mechanical Turk. This will be available at http://ielab.github.io/relevation

References

[1] Bevan Koopman. *Semantic Search as Inference: Applications in Health Informatics.* PhD thesis, QUT, 2014.

[2] L. Goeuriot et al. ShARe/CLEF eHealth Evaluation Lab 2013, Task 3: Information retrieval to address patients' questions when reading clinical reports. In *CLEF*, 2013.

[3] D. Hawking, N. Craswell, P. Bailey, and K. Griffihs. Measuring search engine quality. *Inf. Ret.*, 4(1):33–59, 2001.

[4] D. Lewandowski and S. Sünkler. Designing search engine retrieval effectiveness tests with RAT. *Information Services and Use*, 33(1):53–59, 2013.

---

[1] In the example screenshot the documents are web pages from CLEF eHealth 2013; however, documents could also be plain text documents: Relevation! can be customised by adding document parsers for other formats, e.g. TREC or WARC formats.