

# SCAMP: A Tool for Conducting Interactive Information Retrieval Experiments

Gareth Renaud and Leif Azzopardi  
Information Retrieval Group, School of Computing Science  
University of Glasgow, Glasgow, G12 8QQ  
{0707249r@student, leif@dcs}.gla.ac.uk

## ABSTRACT

Conducting Interactive Information Retrieval (IIR) research is often seen as an arduous and tedious process with a high barrier to entry. This high barrier is due to the overheads in developing and setting up even a simple lab-based IIR experiment. SCAMP (Search ConfigurAtor for experiMenting with PuppyIR) is a web-based tool that we have developed which enables researchers to configure standard IIR experiments. SCAMP provides the infrastructure that handles the major processes within the experimental flow (such as Participant Registration, Consent, Surveys, and the logging and tracking of tasks and participants through the experiment). Consequently, SCAMP reduces the time required to create an experiment and is ideal for undergraduate and masters students who would like to conduct an IIR experiment without extensive development. Furthermore, the tool is extensible. Other features can be easily added to SCAMP to customise the experiments. To evaluate SCAMP, we performed a within-subjects experiment where 48 participants used a web search engine with different search aids (query suggestion, spell correction, etc) to complete various web search tasks – to increase the difficulty of the search tasks certain query terms were banned. We use this evaluation to showcase the different features of the SCAMP system and report on how participants perform under difficult querying conditions.

## Categories and Subject Descriptors

H.5 [Information Interfaces and Presentation]: User Interfaces—*Interaction styles*; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—*Information Filtering*

## General Terms

Information Retrieval, Taboo Search

## Keywords

Information Retrieval, Online Experiments, Configure, Conduct, Analyse

## 1. INTRODUCTION

Interactive Information Retrieval experiments often provide invaluable and deep insights into the retrieval process [6]. However, in contrast to traditional systems-based experiments epitomized by the TREC evaluation methodology [8],

IIR experiments are often: (i) less controlled, (ii) more expensive to conduct, (iii) require more time to undertake, and (iv) are more difficult to replicate, repeat and reproduce. This is because no fixed methodology exists to evaluate the interaction between users and systems. Instead, various methodologies exist and the appropriate one needs to be selected/developed for the IIR experiment [5]. The overheads in designing an IIR experiment also includes the development of usable experimental IR application (i.e. interface and IR system [2]) along with developing software to log, survey and evaluate the system, participants and interactions. In systems-focused research, however, IR Toolkits such as Lemur and Terrier can be easily extended and used within the TREC methodology to quickly obtain results. Such toolkits, along with the TREC methodology, facilitates replication and reproduction, whereas in a live interactive environment it is difficult to reproduce/replicate experiments to the same extent. To some degree this is because the IR applications are seldom made available to others and the methodology undertaken may not be fully explained. The other source of variation comes from the users. Additionally the inclusion of users requires additional effort in terms of gaining ethics approval, recruitment, and performing the experiments. Each of these points contributes to the perception that undertaking user-focused research is difficult, time consuming and expensive, despite the merits of undertaking such research.

While it is not possible to solve all the associated problems with conducting IIR experiments, it may be possible to ease the burden of undertaking such experiments by developing a tool that enables researchers to design and configure IIR experiments. To this end, we have developed SCAMP, a prototype web based tool that enables researchers to:

1. create and configure an IIR experiment,
2. coordinate and conduct the experiment, and
3. monitor the experiment and see an initial rudimentary analysis of the results.

Furthermore, SCAMP provides importing and exporting facilities to share experimental designs and results. In the remainder of this paper, we shall outline the SCAMP tool before describing a simple IIR experiment we conducted using SCAMP to test out the tool.

## 2. RELATED IIR EXPERIMENT SYSTEMS

Designing and developing systems, tools and frameworks to support research is an important, but often undervalued,

part of the research process. This has meant despite many user-focused studies being conducted only ad-hoc tools have been developed in-house to support local IIR research. This is probably because designing a configurable tool is a significant development challenge, interfaces are often highly varied and novel, and with the rapid changes in technology it is difficult to support such a tool. Nonetheless, a few tools have been developed: WiIRE [7], PooDLE IIR Framework [4] and CIRSE [3].

WiIRE (Web Interactive IR Experiments tool) was designed to provide researchers with a configurable tool for creating and running experiments [7]. The tool enabled the configuration of experiments including the provision of questionnaires, tasks and other essential experiment components. The WiIRE framework was tested on the TREC 11 Interactive track in 2003/2004.

Toms *et al* [7], found significant cost savings in the design and running of their experiments. However, WiIRE does not provide data analysis, multi-session experiment participation or self registration. In [7], they mainly focus on reporting the system from the participants' perspective so it is not clear how or how much configuration/coding a researcher would need to perform to use or extend it.

PooDLE is an experiment framework which focuses heavily on logging user interaction [4]. Several loggers are supported, such as UsaProxy [1] and PooDLE is extensible for more to be added. Acknowledged limitations are that there is no interface for easy visualisation of the logging data and PooDLE lacks an intuitive interface for easy configuration, control and monitoring of experiments. PooDLE has been released to the community as an open source project<sup>1</sup> and makes part of the experimental process is made easier but it is not a complete turnkey solution. The configuration capability appears to be there but the authors acknowledge the need for more work in this area to make configuration easier for the researcher.

CIRSE [3] is a very promising framework which supports experiment configuration and data analysis. However the user interface for experiment configuration by researchers is still under development. Logging behaviour and data analysis has a stronger focus than the experiment configuration and participation process.

These systems are all promising attempts at streamlining how IIR experiments are conducted. Unfortunately, WiIRE, PooDLE and CIRSE are either not publicly available or do not provide a complete solution for researchers to create experiments.

### 3. SCAMP OVERVIEW

Here we present SCAMP, an Open Source tool which is included in the PuppyIR<sup>2</sup> framework that is publicly available from SourceForge<sup>3</sup>. The main goals of SCAMP are: (i) to support the creation of common experimental designs and, (ii) to allow researchers to configure and tailor the experiment for their own purposes. The design of SCAMP is similar to that of WiIRE and follows the standard workflow of standard IIR experiments (i.e. participant registration, consent, pre-experiment questionnaire, pre-task questionnaire, task, etc ). A further design goal of SCAMP is to

<sup>1</sup><http://piirexs.sourceforge.net/>

<sup>2</sup><http://www.puppyir.eu>

<sup>3</sup><http://sourceforge.net/projects/puppyir/>

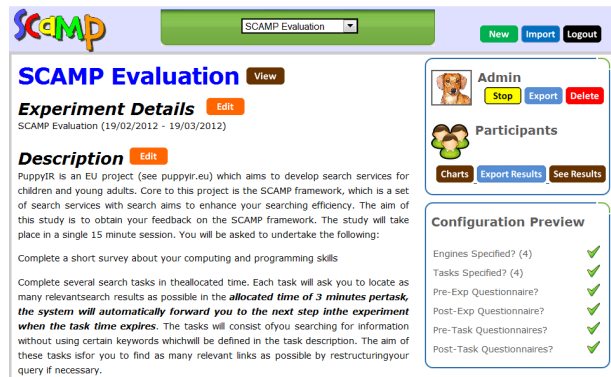


Figure 1: Researcher Dashboard

enable researchers (particularly undergraduates and masters level students) to create and perform simple IIR experiments without any extensive coding (and potentially without coding, just configuring). Thus, the SCAMP tool can be considered as an introductory tool that helps expose students to IIR experiments without excessive overheads.

While there are various types of IIR experiments such as Lab/Task based, Naturalistic, Longitudinal, Cases Studies and Simulations [5], SCAMP currently supports Lab/Task based studies. Furthermore, it currently only supports within-subjects experimental design. While this limits the applicability of the tool, there is sufficient variation to generate interesting and varied IIR experiments.

SCAMP is composed of two applications: (i) The Researcher Dashboard and (ii) The Participant Interface. The first allows the researcher to configure and conduct experiments and to see preliminary results, and the second conducts the experiment.

#### 3.1 Researcher Dashboard

Within the Researcher's Dashboard (see Figure 1), the researcher can create, edit and import/export an existing experiment. To configure and set up an experiment the researcher needs to customize the following components:

**Experiment Details:** The researcher provides a name, the start and end date and a detailed description of the experiment (i.e. Information Sheet and Debriefing Information).

**Consent:** Different consent options can be added along with detailed description to provide participants with consent information. Check boxes are provided to participants to let them agree (or not) to different options.

**Experimental Conditions:** SCAMP assumes that the experimental condition in the experiments is the IR system/engine. Thus the researcher needs to select and configure at least two systems. Since SCAMP uses the PuppyIR framework it can access the engines available with the framework such as Bing, Twitter, YouTube, iTunes, Wikipedia, Picassa, Flickr, and Digg. Each has their own distinct set of tailorable features. Researchers can also specify whether participants are given spelling suggestions and/or related topic suggestions.

**Search Tasks:** The researcher can directly assign a search task to an experimental condition (i.e. system) or specify random assignment for each participant. A

time limit for search tasks can be imposed, and it is possible to specify whether certain query terms are not allowed (or black listed) for the particular task.

**Questionnaires:** Pre/post-experiment and pre/post-task questionnaires can be added. Questionnaire questions can be specified as a short answer (text field), open-ended (text area), options (radio buttons) or Likert Scale. This enables the researcher to design and collect basic information from participants.

Once configured, the experiment can be activated, so that participants can undertake the experiment. The researcher can monitor the activity of participants and their performance/interactions. Once the experiment has been completed the results and data can be exported, along with the experimental setup if the researcher would like to share the design.

## 3.2 Participant Interface

The participant interface provides potential participants with the opportunity to register to undertake experiments. The registration process requires participants to enter a valid email address to which a link is emailed so that their details can be confirmed. Registered users can view the available set of active experiments, and then select it to view the details of the experiment (i.e. read the information sheet). If the experiment is of interest, then they can proceed to the consent page. Here the registered user can elect to participate by checking and agreeing to the consent item options (if participants do not agree to give consent, they cannot participate in the experiment).

In a typical experiment the participant will be asked to answer the pre-experiment questionnaire to collect demographic or other information. Then the participant will be presented with a task: first pre-task questions, then the task itself, followed by post-task questions.

When participants undertake the task they are presented with a standard search interface and can enter queries (if options like search suggestions or spelling suggestions are offered then these are also displayed). Results are returned in a standard list format typical of a web search engine and participants can inspect results (inline) and mark documents they believe are relevant/non-relevant (using thumbs up and thumbs down icons). Participants can examine different pages or results, or re-issue queries. They can continue searching until they complete their task, or in the case of timed tasks if they time out. As previously mentioned, the order of the tasks and conditions is defined in the Researcher's Dashboard. Once all search tasks are complete the post-experiment questionnaire is presented to the participant before the experiment concludes with a debriefing note and thank you message.

## 4. EVALUATION

We evaluated SCAMP in two ways through a usability analysis and by performing an IIR experiment. The usability analysis was performed with an experienced HCI researcher and an undergraduate student (i.e. a likely user of the system).

The researcher found the interface easy to understand and follow, but identified several usability problems (which were subsequently addressed). For example, when spelling suggestions and related search suggestions were displayed they

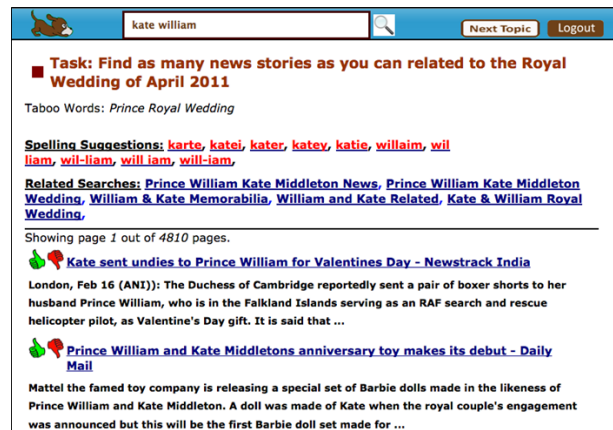


Figure 2: Participant Search Interface

were not clickable, the search box did not display the search query, the cursor did not provide a hover cue and the login, register and password reminder pages displayed inconsistent flow of control. The student also found the interface straightforward and both were able to successfully build a simple IIR experiment. This was aided by the SCAMP assistant which guides the researcher step-by-step through the experiment creation process.

To test the Participant Interface, an experiment was created where we aimed to test how well users could find information when certain query terms were not allowed. The intuition behind the experiment was to simulate query formulation problems i.e. when users are unable to generate queries with terms that are relevant to the topic. Participants were required to find relevant documents without using taboo terms. For example, if the topic was “[country name] independence” then the participants were not allowed to use the terms “[country name], [country name - ish] independence, independent”. The restrictions on query terms increased the difficulty of the topics, requiring participants to use synonyms and related terms to find relevant documents. The experimental condition was the systems where we varied the search engine (Bing): (i) Bing with no search aids, (ii) Bing with spelling suggestions, (iii) Bing with related search suggestions, and (iv) Bing with both spelling and related search suggestions. Four search tasks were created based on high profile news stories (in the country in which the study was undertaken). These search tasks were randomly assigned using a latin square rotation to the different experimental conditions. For each search task/topic, taboo search terms were specified.

Forty-Eight participants undertook the web based experiment of which 28 were male, 19 were female, and one declined to give their gender. A further 16 participants began the experiment, but did not complete it. The experiment was advertised on web based forums and locally. Most participants were from the local country, though other participants were based in various countries throughout the world. In a pre-experiment questionnaire participants were asked how well they could generate synonyms to which the mode was 7 out of 10. However, by the end of experiment, participants rated how well they did on average as 4 out of 10. While the experiment is underway SCAMP provides an overview of the time taken for each participant to complete the experiment (see Figure 3). In addition to this SCAMP

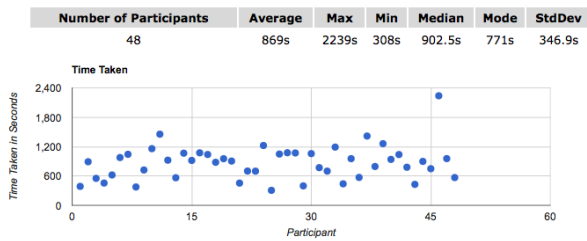


Figure 3: Time taken to complete experiment.

Condition	Avg. Num. Queries	Avg. Num. Relevant Results	Avg. Time (secs)
Bing No Aid	7.35	5.27	138
Bing Spell	7.85	5.5	150
Bing Related	7.93	6.16	144
Bing Both	7.60	6.95	144

Table 1: Usage for each system per task/participant.

provides the researcher with a number of other charts to monitor the performance, such as relevance per engine, relevance per condition, timings, search terms used, number of queries per condition and a summary of question responses. Figure 4, for example, shows the number of relevance judgements that have been made per Engine. Since these charts only provide an initial indication of the outcome of the experiment, SCAMP enables the researcher to export data for further analysis.

## 5. DISCUSSION & FUTURE WORK

In this paper we have introduced SCAMP, a tool for conducting web based IIR experiments. This Open Source Tool can be downloaded and installed via SourceForge (from PuppyIR). The tool enables researchers to create and configure simple within-subjects experiments with a series of different search systems. The idea behind SCAMP was to provide researchers (particularly undergraduates and masters students) with the ability create simple IIR experiments quickly and simply so that they can gain experience and confidence in running such experiments. SCAMP helps achieve this goal and provides a number of other advantages such as enabling web based experiments, inline questionnaires (instead of using third party software or paper-based forms), a

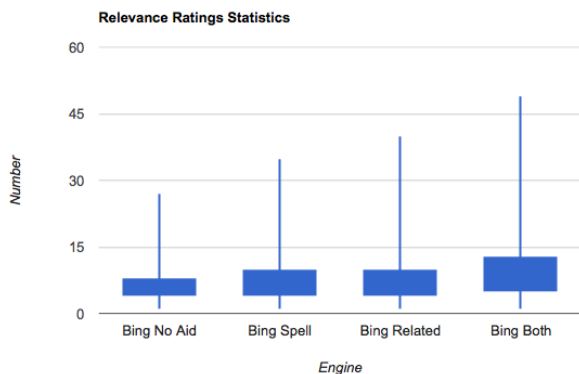


Figure 4: Documents judged relevant by participants for each experimental condition.

number of reusable components for sharing questions, topics and engines, the ability to export and share experimental designs, supports multiple live experiments running as well as simultaneous participation, along with some basic data analysis tools. All of this eliminates much of cost of developing and creating an IIR experiment and lets the researcher focus on the design of the experiment. Since SCAMP is web based participants can undertake the experiment in the lab or remotely.

Currently, the SCAMP tool only supports within-subjects experimental design. Despite this limitation, numerous experiments can be set up as the researcher can select different engines, set up their own topics, questionnaires (and questions), etc. Adding support for other experimental designs such as between-subjects and support for naturalistic experiments is something that we are considering for future development. The system, at present, does not have a reminder system to remind participants to return to complete unfinished experiments. Other features which would enhance SCAMP include support for recruiting and managing participants (i.e. a recall tool to invite participants to finish multi-session experiments or undertake new experiments), integrating an implicit logger (i.e. to capture mouse moves, clicks, etc) as opposed to the current explicit activity logger, greater variety of possible types of survey questions, and also to enable support for multi-researcher accounts (currently there is only one research account). While there are still a number of improvements to be made, SCAMP provides researchers with a simple, but configurable and flexible, tool for creating and conducting IIR experiments.

**Acknowledgements:** PuppyIR is funded by the EC’s FP7 2007-2013 under grant agreement no. 231507.

## 6. REFERENCES

- [1] R. Atterer, M. Wnuk, and A. Schmidt. Knowing the user’s every move. *In 15th International World Wide Web Conference, Edinburgh, Scotland, 2006.*
- [2] L. Azzopardi. Usage based effectiveness measures: monitoring application performance in information retrieval. *In Proceedings of the 18th ACM CIKM ’09*, pages 631–640, 2009.
- [3] R. Bierig, M. Cole, J. Gwizdka, N. Belkin, J. Liu, C. Liu, J. Zhang, and X. Zhang. A user-centered experiment and logging framework for iir. *Understanding the User Workshop at SIGIR’09*, 2009.
- [4] R. Bierig, J. Gwizdka, and M. Cole. A user-centered experiment and logging framework for IIR. *Proceedings of the SIGIR 2009 Workshop on Understanding the User*, pages 8–11, 2009.
- [5] D. Kelly. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval*, 3:1–224, 2009.
- [6] I. Ruthven. Interactive information retrieval. *Annual Review of Information Science and Technology (ARIST)*, 42:43–92, 2008.
- [7] E. G. Toms, L. Freund, and C. Li. WiRE: the web interactive retrieval experimentation system prototype. *Inf. Process. Manage.*, 40(4):655–675, May 2004.
- [8] E. M. Voorhees and D. Harman. *TREC: Experiment and Evaluation in Information Retrieval*. MIT Press, 2005.





## Search Configurator for experimenting with PuppyIR

### Abstract

SCAMP aims to provide an infrastructure for creating and running experiments, creating and conducting questionnaires, some rudimentary data analysis, export experimental designs and to export data for external statistical analysis.

SCAMP is a web based tool and allows participants from anywhere in the world to participate in experiments.

### Aims

Scamp aims are to:

1. Provide an intuitive experiment configuration interface
2. Allow multi-session participation in multiple experiments from anywhere
3. Facilitate data visualisation and analysis

### SCAMP Online Demo

<http://www.dcs.gla.ac.uk/access/scamp>

Search results for 'kate william' showing a task: 'Find as many news stories as you can related to the Royal Wedding of April 2011'. It includes spelling suggestions, related searches, and news snippets such as 'Kate sent undies to Prince William for Valentines Day' and 'Prince William and Kate Middletons anniversary toy makes its debut'.

### SCAMP Researcher Dashboard

The dashboard shows 'SCAMP Evaluation' details, including a description of the experiment tasks and a configuration preview table.

Configuration Preview	
Engines Specified? (4)	✓
Tasks Specified? (4)	✓
Pre-Exp Questionnaire?	✓
Post-Exp Questionnaire?	✓
Pre-Task Questionnaires?	✓
Post-Task Questionnaires?	✓

### Experiment configuration and information

### Experimenter Interface

Experiments are created by researchers using the dashboard

1. A researcher creates/configures experiments
2. The PuppyIR framework is facilitated for API configuration
3. Scamp provides data exporting and visualisation

### Participant Interface

Subjects use the participant side of Scamp

1. Scamp collects demographics and consent from subjects
2. Subjects perform tasks and answer questionnaires
3. Scamp debriefs subjects

### Services Used

PuppyIR<sup>2</sup>: is an open-source framework designed to be used in the creation of new web search services for children. PuppyIR handles the retrieval of results from various services (like YouTube) and presenting these results. Scamp, via PuppyIR, uses the Google YouTube video API and YouTube's filtering system (to remove adult content) is used in addition to PuppyIR filters. In addition Ockham is used for spelling suggestions and Bing provides the related searches.