ECIR Draft
Paper Guidelines

BCS Information Retrieval Specialist group

8th November 2006

Version 1.0

Preface

On the 20th of June, 2006 a workshop on the European Conference in Information Retrieval was held which aimed to develop a set of guidelines for authors and reviewers of ECIR papers. These draft guidelines have been complied based upon the presentations from the workshop, and have been extracted from the full workshop report, which is available to download from the BCS-IRSG website (http://irsg.bcs.org). We hope that this document will serve as the basis for helping authors wishing to submit to ECIR (especially student and first time authors), along with aiding reviewers in the refereeing process. These guidelines, are exactly, that guidelines, and authors and referees need to use their common sense, experience and discretion when using and interpreting these guidelines.

If you have any questions, comments, suggestions or contributions for the draft guidelines please contact the BCS-IRSG Chair, Leif Azzopardi (leif at dcs.gla.ac.uk), the ECIR Coordinator, Iadh Ounis (ounis as dcs.gla.ac.uk) or the BCS-IRSG Secretary, Andy MacFarlane (andym at soi.city.ac.uk).

The BCS-IRSG would like to thank all participants and contributors, especially Iadh Ounis (University of Glasgow) and Andrew MacFarlane (City University), Keith van Rijsbergen (University of Glasgow), Ian Ruthven (University of Strathclyde), Ayse Goker (City University), Mounia Lalmas (Queen Mary University, London), Gianni Amati (FUB) and David Losada (Universidad de Santiago de Compostela). The BCS-IRSG would also like to thank the Department of Computing Science at the University of Glasgow for hosting the event.

How to write a good ECIR paper

Introduction

The purpose of this section is to provide a set of guidelines for authors wishing to submit to ECIR. The goal is to provide an overview of the different types of papers that are acceptable and what is required given that type. We provide a guide for the following areas: theoretical, experimental/comparisons, People in IR, applications, conceptual, and evaluation and performance measures. The guide is also intended to aid reviewers of ECIR papers, as well to provide an indication of the expected level of work and the types of acceptable papers.

Properties of a Good IR paper

A good IR paper or any scientific paper for that matter should engage the reader and succinctly convey an original idea. The contribution to knowledge should be expressed as clearly as possible to facilitate understanding and provide the reader with a better insight. For instance, the paper should aim to provoke such reactions from the reader along the lines of, "I now better understand this issue.", "It works!"). Or enables the reader to draw connections between other bodies of research ("I'll see how this fits with my XYZ approach", "It can also be applied to ABC!"). Or invokes the reader's imagination or changes their perspective ("This paper has got me thinking!", "I haven't seen this from this point of view before."). A great IR paper will obtain a response from the reader like "Why didn't I write this paper!". Of course, not all papers invoke such responses, for various reasons. Here, we focus on some of the main genres of papers that are encouraged to be submitted to ECIR, and describe some of the qualities that they should possess.

Please note that these are guidelines and authors and reviewers should use them in the way intended; as guidelines to help the process of writing and reviewing. Consequently, the author and referee should use their experience, common sense and discretion when interpreting and using these guidelines.

Theoretical Papers

What is a theoretical paper?

A theoretical paper proposes a theory for Information Retrieval (or some phenomena within the domain). A theoretical paper should present a supposition or system of ideas intended to explain some phenomena within IR. It should be based on general principles independent of the phenomena to be explained (i.e. Darwin's Theory of Evolution). It could provide a set of principles on which the practise of an activity is based (i.e. a theory of information seeking behaviour). Or it could present an idea used to account for a situation or justify a course of action. Consequently, this does not necessarily imply that the theory is grounded in mathematics or some other formalism, which is common misconception about theoretical papers in IR. Instead, a theoretical paper may also be discursive in nature, providing arguments and reasoning through the discourse.

What makes a good theoretical paper?

First, the paper must go beyond the existing theory already present in the literature – and thus fulfil the originality criteria. In order to convince the reader that this is the case, there should be links to older theory to provide the context of the paper. The relationship between the old and the new should be related and explained.

Second, it is important for a theoretical paper in IR to provide the necessary contextualisation of the theory within IR. That is, what is the relevance of this theory to IR? Consequently, the generic application of a machine learning approach, for example, is not relevant. The burden is on the writer to exemplify the link between the theory and the practise, given the domain.

Third, the clarity of the presentation is very important, because the emphasis of the paper is to present an account for a phenomenon. Consequently, the arguments presented need to be clear and justified. One way of ensuring clarity is to provide illustrative and practical examples to aid the reader's understanding.

Forth, a theoretical paper aims to link theory with practise; once a theory is presented, the inevitable question arises; does it work in practise? However, "proof" that a theory holds is not a necessary requirement for a theoretical paper to be acceptable, as it is not always possible for a theory to be put forward and for it to be tested to the nth degree. There are various reasons for this; the work is in its early stages; the machinery doesn't exist for it to be tested; etc.

In such cases when experimental work can not be provided to ensure that the paper is acceptable, there are other criteria that the paper should meet. A discussion should be included about the testability of the theory presented, comments on whether it can be falsified, its tractability, how the theory could be tested in practise, its relationship with experimentation, and whether it is possible to implement or not. Addressing such issues is paramount to papers, which present novel/new theory.

However, there are cases when the theory presented is an extension to the existing theory. In this case, where the theory has been tested previously, it is necessary to provide some experimental work in order to show that this extension is actually significant, useful, successful etc. Re-stated, delta theory papers should provide some empirical testing. On the point of significance, a theoretical paper should also discuss what would constitute a significant result and how to quantify this.

Experimental/Comparisons Paper

What is an Experimental paper?

An experimental paper compares one or more competing theories/techniques within Information Retrieval. An experimental paper should contain the context of the study, a clear statement of the problem addressed, and present clear research hypotheses.

What makes a good experimental paper?

The paper should make an original contribution to IR and state clearly what exactly is new with respect to previous work. Consequently, a good set of references should be included to link prior work; and should include those approaches, which can be used as a baseline. An experimental paper should use publicly available and (preferably) standard test collections. For instance, the generally accepted collections for empirical evaluation are those provided by TREC, but also include collections from other common evaluation forums such as CLEF, INEX, etc. The use of older collections such as Medline, Cranfield, CACM, NPL, etc are now considered too small to be acceptable. Experimental papers that use parts of test collections or subset of topics are generally considered unacceptable, unless accompanied by a reasonable justification (see below).

The use of a non-standard test collection can be acceptable if it is publicly available, and representative or diverse enough to warrant reasonable conclusions. Note that the requirement of test collections being publicly available is to ensure that the experiments performed within the paper can be reproduced. Consequently, a good paper will ensure that the data is available to enable replication, verification, and/or reproducibility of the work. If the data used in experiments is not publicly available then the findings, reports and conclusions drawn should not be based on such data. For instance, such experimental analysis may complement the findings shown on publicly available data collections. Or, it could be used to illustrate or illuminate the findings but are not pinnacle to the contribution.

An experimental paper should justify the data collection(s) and analysis methods used. Depending on the retrieval task, the paper should use an appropriate test collection (generally the most recent one), all its associated topics and assessments coupled with a suitable analysis method. In particular, the use of a non-standard test collection should be justified and ensure that there is access to the collection (either, through the author, on their website, etc). A good experimental paper should use more than one test collection (if available) to provide more evidence for the hypotheses presented and show how generalisable the techniques examined are.

An experimental paper should use appropriate statistical or qualitative methods and report appropriate and standard evaluation measures. However, simply performing and reporting significance test and so forth is not sufficient without further explanation of that significance (see below).

Importantly, an experimental paper should use appropriate state-of-the-art baseline(s) to convince the reader that the proposed technique is superior or not. The characteristics of a good baseline include that it provides strong retrieval performance, is well-established and robust across different collections. Other baselines may include the best performing runs at TREC, for instance. Though, they should not be used as a sole indicator (as such runs have not been optimized/tuned given the collection). If tuning or optimizations of the techniques or models are required then this luxury should also be afforded to the competing baseline

models. If the experimental paper proposes a technique, which is computationally expensive, this inefficiency should be acknowledged and discussed.

Finally, an experimental paper should indicate the significance of the results and conclusions made with respect to the practice and/or theory of IR. By reflecting on the methods, datasets and results (or combination there of) the significance of the work should be provided to explicitly and precisely describe how the work is better/useful/interesting/etc with respect to the current start-of-the-art. An excellent paper will provide insights into why it is succeeding or failing and also discuss how generalisable the results are. A key point, here, is that a paper in this genre is still acceptable even if the model does not perform as expected, so long as this insight is provided. Conversely, a paper presenting outstanding results should not be simply accepted on the basis of the superior results, if there is no rationale and understanding to how these results were achieved.

"People in IR" - User studies/Interfaces papers

What is a "People in IR" paper?

"People in IR" type papers cover a variety of research within Information Retrieval; the distinguishing feature of these papers is that they involve humans as a major component in the system, experiment or investigative study being described.

Broadly, there are two main types of "People in IR" papers: (1) those based on laboratory IR investigations – these are similar to experimental papers but with the involvement of humans, and (2) those investigating information seeking and behaviour. The first type, generally referred to as Interactive IR, includes evaluation of novel interfaces, and interaction work, including user modelling and predictive or adaptive technologies (with people involved in the evaluation or data collection). The second type, dealing with what is commonly called Information Seeking or Information Behaviour, is more concerned with the information needs and search behaviour of individuals or distinct groups of people. Papers in the Information Seeking area may be more discourse-based than is normally seen at ECIR. Both types of research are, however, welcomed at ECIR.

There are core approaches and techniques to facilitate research in the area of interactive IR and Information Seeking. However, there are a few fixed methodologies. Due to to the variety of research within this genre, a "people in IR" paper usually describes a novel methodology specifically created for an individual investigation. This methodology – a coherent set of decisions and investigative components – and the reasons behind the methodology will require explanation within the paper. This is in contrast to other branches of IR where we have standard methodologies, metrics and tools such as test collections that are commonly accepted within the discipline and require fewer introductions.

What makes a good "People in IR" paper?

Any good "People in IR" paper should provide a coherent narrative to describe the research questions motivating the research, the methodology to investigate these questions – including the design decisions behind any novel system or interface development – the relevant results obtained and the implications for future research.

A laboratory based IR paper is, in many respects, similar to experimental papers and the same general guidelines apply. The introduction of people within the studies however introduces some further issues. In particular, the paper should describe the people involved and why their particular characteristics, such as search experience, might influence the results obtained. Similarly, the paper needs to describe the components of the study, such as the source of search tasks, any baseline systems, and instructions given to participants in the study and how these aspects relate to the research questions and results obtained. It is very important that a paper shows these connections to convince the reader that the experimental set-up is not unrealistic or biased.

An Information Seeking paper should also explain the methodologies chosen to investigate the research goals of the paper and, where appropriate, discuss alternate methodologies that could also have been followed. A good Information Seeking paper will investigate search phenomena in depth rather than just reporting or describing the study and basic results. Rather, a good paper will seek to investigate the reasons for the results and will present and analysis of the implication of the findings for IR research.

For both Interactive IR and Information Seeking papers, it is important that the evaluation and methodology should be appropriate given the research hypotheses and

objectives. Common criticisms of such papers are that there are not enough participants, user groups, tasks, or baselines. However, such criticisms should only be made with respect to the methodology and research questions presented. For example, an evaluation of an interface that is intended to be used by a wide group of searchers for all search tasks will require more participants and more search tasks before we can accept that the results obtained are meaningful. A more qualitative investigation on, for example, children's uptake of new search technologies, may require fewer participants because each participant will be analysed in more depth. Consequently it is important to justify such design choices and also to explicitly acknowledge any limitations of the study and how such limitations might affect the outcomes of the study.

In both types of study, many results will be obtained, all of which cannot be presented in one paper. Therefore it is necessary to select results rather than overwhelm the reader with as many as possible. To avoid appearing to be overly selective (i.e. only presenting results that are positive with respect to some existing hypotheses) it is better to concentrate on a smaller number of related results and research questions and investigating them in more depth. Unexpected or surprising results are worth including as is qualitative information from any participants in the study. Qualitative information helps contextualise the quantitative results and helps the reader understand the experience of the human participants.

What makes a poor "People in IR" paper?

A weaker paper in both Interactive IR and Information Seeking can suffer from a number of problems that make it hard for the reader to appreciate and follow the research being presented. A common flaw, and one that will kill most papers, is poor exposition of the research itself – not explaining why the research is being carried out, how the study was constructed, which individual results were selected for presentation, and how the research links with other research in the area. A particular weakness for Information Seeking papers, when submitted to a conference such as ECIR, is not to discuss the implications of the work for the general field of Information Retrieval – how might the research being presented change the way IR systems or interfaces are designed or evaluated?

Methodologies for evaluation or investigation of behaviour are usually complex and we cannot explain every detailed aspect of the methodology within the limits of a conference paper. However, it is important to provide sufficient detail so that the reader can follow the study being described but also the thinking behind the design of the study.

A second and common problem, particularly for Interactive IR papers, is to simply present the quantitative results rather than describing them and their importance. In a test collection evaluation it *may* be acceptable simply to show statistically significant results between standard evaluation metrics in order to convince the reader of the benefit of one algorithm over another. In an Interactive IR paper, this is usually not enough; rather the results have to be explained to the reader. If searchers run more queries on a novel interface than on a baseline, for example, the author should explain why this behaviour might have occurred with reference to the design principles of the interfaces and preferably to the other results obtained within the study to present a full understanding of the importance of the result.

The most common flaw with Information Seeking papers is simply to present the author's experience of running the study with no attempts to validate their findings (e.g. by using a mixture of elicitation methods), no attempt to discuss the implications of the research and no attempt to relate the study to existing work in the literature.

Applications and System Prototype papers

What is an applications paper?

There are four main types of application papers:

(1) Positioning papers;
(2) Technical papers;
(3) Demo papers; and
(4) Test and evaluation papers.

A positioning paper details the motivation and background for an application. A technical paper details the description of the architecture, individual components, algorithms, integration of components, etc. A demo paper describes the system in the paper and is usually coupled with a demonstration of the application. A test and evaluation paper reports empirical results from the testing of an application.

An applications paper, therefore, may report different stages or phases of a research project and an application prototype and its development; such phases include requirements and design analysis, Prototyping and implementation, testing and evaluation, and dissemination. An applications paper may iterate across the above phases to produce papers of type 1 to 4 respectively. Obviously, the expected contribution and impact of the paper will/should grow as later phases are reported.

What makes a good Applications papers?

An applications paper should include an explicit system description and take the reader through a user experience or scenario, if applicable, providing examples such as a walk through of the system and the iterations.

It is important that an applications paper contextualises where in the timeline, the application is at in the series of phases. Further, this contextualisation should also include how the work relates to the system as a whole. To facilitate this, it is important that related work is cited in the relevant disciplines. As a consequence, the reader can fully appreciate and contextualise the work, and the contribution's references to any prior work should be included.

A positioning paper should present a thorough and comprehensive background along with a detailed motivation for the application. The uniqueness of the solution/application should be explained and the justification for how this position was arrived should be provided.

A demo paper should provide a description of the system and how the system would be experienced by the user. Since it is a demo, there is an expectation that presentation will contain a demonstration of the application/prototype system; i.e. the system description is of the demo, rather than what is to be built, and the system should be working such that reasonable feedback can be given by the audience. The system description should include the science and motivation behind the application to justify why the application is novel and warrants demonstration. Along with the system description should be a technical specification stating the configuration, hardware and other requirements that the application requires. This should be done separately from the system description to avoid oscillating between motivation/science and the technical aspects.

In a Test and evaluate paper, a clear distinction should be made between the testing of the system (through running experiments, etc) and the evaluation and analysis of the experiments. During the test process, the inclusion of a functionality check should be included to detail what is operational in the application and what is not, and to specify any other limitations relevant to the experimentation. As with user studies papers, a good applications paper of this kind, will describe the experimental conditions under which the user testing took place; for instance pointing out whether real users were involved or whether it was pilot tested on colleagues.

What makes a poor applications paper?

An applications paper that presents an idea (new or old) but no evidence or explanation of its perceived need or uniqueness is not acceptable. A paper, which is a technical push of some technology, is not acceptable, unless a clear requirement for the specific technical improvement is shown and justified, or if a novel vision is presented where the technical push would be appropriate. Application papers should provide sufficient evidence to motivate or justify their arguments. Consequently papers which rely on hearsay, word of mouth, or practise which has become a de-facto standard, do not provide adequate justification.

Conceptual papers

What is a conceptual paper?

A conceptual paper presents concepts dealing with or relating to IR, where a new perspective is obtained or formulated by the combination of a group or class of objects. For instance, the identification of trends or patterns, which occur in IR, where the contribution to knowledge is the definition of the concepts and their relationships to the IR process. An example of a conceptual paper is one that defines a new way to characterise the notion of relevance in IR, and shows how this new way links to other, and also what it brings to what was there before.

In general, a conceptual paper is likely to be discursive. This is to our point of view what makes a conceptual paper different to a formal/theoretical paper, although the boundaries here can be very fuzzy. Also a conceptual paper requires a very deep understanding of the problem or issue being addressed, investigated, studied etc and as such we would say that such papers are difficult to write.

What makes a good conceptual paper?

A good conceptual paper should state and formally define all the concepts introduced and their relationships/interactions. This can be thought of as the conceptual development [AMSR].

The conceptual development should strive to maintain consistency in the level of abstraction and the unit of analysis. Also, when using multidimensional concepts (for example in INEX, relevance is defined as a two-dimension multi-graded concept), all the relevant dimensions should be clearly explained. The underlying assumptions of the concepts should also be explicitly stated (e.g., conceptual paradigm), along with the boundaries or limitations of the conceptual development. Consequently, a good conceptual paper will also explain for whom and under what conditions the proposed conceptual development is appropriate or inappropriate.

Not only must conceptual paper justify the conceptual development, but it should also show why this is a viable way to do so, as opposed to other ways. Hence, a good conceptual paper establishes a clear link between the proposed conceptual developments and previous research (including alternative perspectives) and explains how the conceptual development improves upon existing conceptualizations. This is very important, and requires a strong knowledge of the problem/area by the author.

An important part of the contribution made by a conceptual paper is how the proposed conceptual development changes our understanding of the theory and practice in IR. In other words, a conceptual paper should spend considerable effort establishing the implications of the developed concepts, which are being proposed.

Similar to theoretical papers, a conceptual paper should make certain that the conceptual development can ultimately be tested and measured empirically.

A bad conceptual paper is one that contains lots of (eventually) interesting concepts, whether novel or not, but with very little connection to the IR problems being tackled. Defining concepts and their relationships, however elegant, is not enough. We must make sure that we relate the proposed conceptual development to the problems being addressed, and how they relate to state-of-the-art work.

Evaluation and Measurement papers

[To be added]

What is an Evaluation paper?

What makes a good Evaluation paper?

Taxonomy of research papers as applied to IR

According to Parberry [parberry94], research papers can be broken in six main categories: Breakthrough, ground-breaking, progress, re-prise, tinkering, debugging, and survey. Here are the categories with respect to Information Retrieval.

Breakthrough

This is typically an open problem that has persisted for some time, which has received a considerable amount of research effort dedicated to solving that problem. (e.g. dealing and handling context, evaluation, introduction of pioneering models, i.e. language modeling [ponte98lm], BIM [robertson76bim], etc.)

Ground-Breaking

A paper would fall into this category if it opens up a field within IR that is not well explored or understood, and provides a firm foundation for further research. (Examples, vector-space model [salton68vsm], inference networks [turtle90in], non-classical logics [vanrijsbergen86nc], etc.)

Progress

This type of paper would solve or address a new or recent open problem, typically limited to that particular context. (Apparently, most papers are of this variety.)

Reprise

A paper of this variety would provide superior evidence (possibly contradictory evidence) of a previous result. It is important to ensure that there is both elegance and insight obtained by the paper. This could entail the use of a better experimental design to test the hypothesis, which results in more conclusive evidence, an analysis that is more thorough that the previous, makes more illuminating connections than the past work. (Experiments contesting the cluster hypothesis, would tend to be of this type, for instance)

Tinkering

Such papers are only really of merit if the extension of known results is provided through a more careful and detailed analysis, but in a non-obvious way. (As opposed to incremental research, tweaking the algorithm for performance increases etc)

Debugging

Such papers would elucidate and then repair a previously undiscovered flaw in previously published work.

Survey

A paper in this category would unify the particular area of a specialized subject within IR with consistent notations, terminology, etc, often piecing together results from many disparate sources. For instance, Hawkings' paper on Enterprise Search [hawking04es] or Callan's paper on distributed IR [callan00dir].

References

[Parberry94] Ian Parberry. *A guide for new referees in theoretical computer science*. Information and Computation, 112(1), 1994.

[AMSR] http://www.amsreview.org/submit.htm. Academy of Marketing Science Review, Submission Guidelines.

[ponte98lm] Ponte, J. M. and Croft, W. B., *A Language Modeling Approach to Information Retrieval*, In the Proceedings of the Twenty First Annual International ACM-SIGIR, 1998.

[robertson76bim] Robertson, S. E. and Sparck-Jones, K., *Relevance weighting of search terms*, Journal of the American Society for Information Science, Vol, 27, 1976

[turtle90in] Turtle, H. and Croft, W. B., *Inference Network for Document Retrieval*, In the Proceedings of the 13th Annual International ACM SIGIR, 1990

[vanrijsbergen86nc], van Rijsbergen, C. J., *A non-classical logic for information retrieval*, The Computer Journal, Vol. 29(6), 1986

[salton68vsm], Salton, G. and Lesk, M. E., *Computer Evaluation of indexing and text processing*, Journal of the ACM, Vol .15(1), 1968

[callan00dir], Callan, J., Advances in information retrieval, *Chapter5: Distributed Information Retrieval*, 2000

[hawking04es] Hawking, D., *Challenges in enterprise search*, In the Proceedings of the fifteenth Australasian database conference, 2004