

Fabio Crestani  
Ian Ruthven (Eds.)

LNCS 3507

# Context: Nature, Impact, and Role

5th International Conference on Conceptions  
of Library and Information Sciences, CoLIS 2005  
Glasgow, UK, June 2005, Proceedings



Springer

250.1-55  
C746  
2005

Fabio Crestani Ian Ruthven (Eds.)

# Context: Nature, Impact, and Role

5th International Conference on Conceptions  
of Library and Information Sciences, CoLIS 2005  
Glasgow, UK, June 4-8, 2005  
Proceedings



E200501353



Springer

Volume Editors

Fabio Crestani

Ian Ruthven

University of Strathclyde

Department of Computer and Information Sciences

Livingstone Tower, 26 Richmond Street, Glasgow G1 1XH, UK

E-mail: {f.crestani, ian.ruthven}@cis.strath.ac.uk

Library of Congress Control Number: 2005926658

CR Subject Classification (1998): H.3, H.2, F.2.2, H.4, H.5.4, I.6, I.2

ISSN 0302-9743

ISBN-10 3-540-26178-8 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-26178-0 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin Heidelberg 2005

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India  
Printed on acid-free paper SPIN: 11495222 06/3142 5 4 3 2 1 0

*Commenced Publication in 1973*

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

## Editorial Board

David Hutchison

*Lancaster University, UK*

Takeo Kanade

*Carnegie Mellon University, Pittsburgh, PA, USA*

Josef Kittler

*University of Surrey, Guildford, UK*

Jon M. Kleinberg

*Cornell University, Ithaca, NY, USA*

Friedemann Mattern

*ETH Zurich, Switzerland*

John C. Mitchell

*Stanford University, CA, USA*

Moni Naor

*Weizmann Institute of Science, Rehovot, Israel*

Oscar Nierstrasz

*University of Bern, Switzerland*

C. Pandu Rangan

*Indian Institute of Technology, Madras, India*

Bernhard Steffen

*University of Dortmund, Germany*

Madhu Sudan

*Massachusetts Institute of Technology, MA, USA*

Demetri Terzopoulos

*New York University, NY, USA*

Doug Tygar

*University of California, Berkeley, CA, USA*

Moshe Y. Vardi

*Rice University, Houston, TX, USA*

Gerhard Weikum

*Max-Planck Institute of Computer Science, Saarbruecken, Germany*

## Preface

CoLIS 5 was the fifth in the series of international conferences whose general aim is to provide a broad forum for critically exploring and analyzing research in areas such as computer science, information science and library science. CoLIS examines the historical, theoretical, empirical and technical issues relating to our understanding and use of information, promoting an interdisciplinary approach to research. CoLIS seeks to provide a broad platform for the examination of context as it relates to our theoretical, empirical and technical development of information-centered disciplines.

The theme for CoLIS 5 was the nature, impact and role of context within information-centered research. Context is a complex, dynamic and multi-dimensional concept that influences both humans and machines: how they behave individually and how they interact with each other. In CoLIS 5 we took an interdisciplinary approach to the issue of context to help us understand and the theoretical approaches to modelling and understanding context, incorporate contextual reasoning within technology, and develop a shared framework for promoting the exploration of context.

The Organizing Committee would like to thank all the authors who submitted their work for consideration and the participants of CoLIS 5 for making the event a great success. Special thanks are due to the members of the Program Committee who worked very hard to ensure the timely review of all the submitted manuscripts, and to the invited speakers: Prof. David Blair, University of Michigan, Business School, USA and Prof. Elisabeth Davenport, Napier University, School of Computing, UK. We also thank the sponsoring institutions, EPSRC, the Kelvin Institute, the BCS-IRSG and the University of Strathclyde, for their generous financial support of the conference, and Glasgow City Council for its civic hospitality.

Thanks are also due to the editorial staff at Springer for their agreement to publish the conference proceedings as part of the Lecture Notes in Computer Science series.

Finally thanks are due to the local team of student volunteers (Mark Baillie, Heather Du, David Elswailer, Emma Nicol, Fabio Simeoni, Simon Sweeney and Murat Yakici), secretaries (Linda Hunter and Carol-Ann Seath), and the information officer (Paul Smith) whose efforts ensured the smooth organization and running of the conference.

June 2005

Fabio Crestani  
Ian Ruthven

# Organization

## Organizing Institutions

CoLIS 5 was organized by the Department of Computer and Information Sciences of the University of Strathclyde, Glasgow, UK.

## Organization

Conference Chair: Fabio Crestani, University of Strathclyde, UK

Program Chair: Ian Ruthven, University of Strathclyde, UK

Workshops Chair: Jonathan Furner, University of California, Los Angeles, USA

Tutorials Chair: Monica Landoni, University of Strathclyde, UK

Doctoral Forum Chair: Pia Borlund, Royal School of Library and Information Science, Denmark

Advisory Committee: Peter Ingwersen, Royal School of Library and Information Science, Denmark, and Pertti Vakkari, University of Tampere, Finland

## Sponsoring Institutions

Engineering and Physical Sciences Research Council (EPSRC), UK

University of Strathclyde, Glasgow, UK

Kelvin Institute, Glasgow, UK

British Computer Society, Information Retrieval Specialist Group

Glasgow City Council, Glasgow, UK

## Program Committee

Alan Smeaton, Dublin City University, Ireland

Andreas Rauber, Vienna University of Technology, Austria

Andrew MacFarlane, City University of London, UK

Anastios Tombros, Queen Mary University of London, UK

Bernard Jansen, Pennsylvania State University, USA

Birger Hjørland, Royal School of Library and Information Science, Denmark

Birger Larsen, Royal School of Library and Information Science, Denmark

C. J. van Rijsbergen, University of Glasgow, UK

Christine Borgman, University of California, Los Angeles, USA

## VIII Organization

Dagobert Soergel, University of Maryland, College Park, USA  
David Bawden, City University, London, UK  
David Hendry, University of Washington, USA  
Diane Kelly, University of North Carolina, USA  
Dietmar Wolfram, University of Wisconsin, Milwaukee, USA  
Douglas Tudhope, University of Glamorgan, UK  
Edie Rasmussen, University of British Columbia, Canada  
Efthimis Efthimiadis, University of Washington, USA  
Erica Cosijn, University of Pretoria, South Africa  
Forbes Gibb, University of Strathclyde, UK  
Gabiella Pasi, National Council of Research, Italy  
Gareth Jones, Dublin City University, Ireland  
Gobinda Chowdhury, University of Strathclyde, UK  
Harry Bruce, University of Washington, Seattle, USA  
Ian Cornelius, University College Dublin, Ireland  
Irene Wormell, Swedish School of Library and Information Studies, Sweden  
Janna Kekäläinen, University of Tampere, Finland  
Jesper Schneider, Royal School of Library and Information Science, Denmark  
Joemon Jose, University of Glasgow, UK  
Josiane Mothe, Paul Sabatier University, France  
Marcia J. Bates, University of California, Los Angeles, USA  
Maristella Agosti, University of Padova, Italy  
Mark Sanderson, University of Sheffield, UK  
Michael Buckland, University of California, Berkeley, USA  
Michael Thelwall, University of Wolverhampton, UK  
Morten Hertzum, Roskilde University, Denmark  
Nils Pharo, Oslo University College, Norway  
Paul Solomon, University of North Carolina, USA  
Pertti Vakkari, University of Tampere, Finland  
Peter Bruza, Distributed Systems Technology Centre, Australia  
Peter Ingwersen, Royal School of Library and Information Science, Denmark  
Preben Hansen, Swedish Institute of Computer Science, Sweden  
Ragnar Nordlie, Oslo University College, Norway  
Raya Fidel, University of Washington, USA  
Ryen White, University of Maryland, College Park, USA  
Sanna Talja, University of Tampere, Finland  
Sharon McDonald, University of Sunderland, UK  
Stefan Rüger, Imperial College London, UK  
Stefano Mizzaro, University of Udine, Italy  
Theo Bothma, University of Pretoria, South Africa

## Additional Reviewers

João Magalhães, Imperial College London, UK  
Peter Howarth, Imperial College London, UK

## Previous Venues of CoLIS

CoLIS 1 was held in 1991 at the University of Tampere, Tampere, Finland.

CoLIS 2 was held in 1996 at the Royal School of Librarianship, Copenhagen, Denmark.

CoLIS 3 was held in 1999 at the Inter-university Centre, Dubrovnik, Croatia.

CoLIS 4 was held in 2002 at the University of Washington, Seattle, USA.



# Table of Contents

## Invited Papers

Wittgenstein, Language and Information: "Back to the Rough Ground!" <i>David C. Blair</i> .....	1
Text, Co-text, Context and the Documentary Continuum <i>Elisabeth Davenport</i> .....	5

## Representing Context

The Sense of Information: Understanding the Cognitive Conditional Information Concept in Relation to Information Acquisition <i>Peter Ingwersen, Kalervo Järvelin</i> .....	7
Practical Implications of Handling Multiple Contexts in the Principle of Polyrepresentation <i>Birger Larsen</i> .....	20
Information Sharing and Timing: Findings from Two Finnish Organizations <i>Gunilla Widén-Wulff, Elisabeth Davenport</i> .....	32

## Context and Relevance in Information Seeking

Contexts of Relevance for Information Retrieval System Design <i>Erica Cosijn, Theo Bothma</i> .....	47
Searching for Relevance in the Relevance of Search <i>Elaine G. Toms, Heather L. O'Brien, Rick Kopak, Luanne Freund</i> ...	59
Information Searching Behaviour: Between Two Principles <i>Nick Buzikashvili</i> .....	79

## Context and Information

Bradford's Law of Scattering: Ambiguities in the Concept of "Subject" <i>Birger Hjørland, Jeppe Nicolaisen</i> .....	96
The Instrumentality of Information Needs and Relevance <i>Olof Sundin, Jenny Johannisson</i> .....	107

Lifeworld and Meaning - Information in Relation to Context	
<i>Janne Backlund</i> .....	119

## Contextualised Information Seeking

Personometrics: Mapping and Visualizing Communication Patterns in R & D Projects	
<i>Morten Skovvang, Mikael K. Elbæk, Morten Hertzum</i> .....	141

Annotations as Context for Searching Documents	
<i>Maristella Agosti, Nicola Ferro</i> .....	155

Conceptual Indexing Based on Document Content Representation	
<i>Mustapha Baziz, Mohand Boughanem, Nathalie Aussenac-Gilles</i> .....	171

## Agendas for Context

What's the Deal with the Web/Blogs/the Next Big Technology: A Key Role for Information Science in e-Social Science Research?	
<i>Mike Thelwall, Paul Wouters</i> .....	187

Assessing the Roles That a Small Specialist Library Plays to Guide the Development of a Hybrid Digital Library	
<i>Richard Butterworth, Veronica Davis Perkins</i> .....	200

Power Is Information: South Africa's Promotion of Access to Information Act in Context	
<i>Archie L. Dick</i> .....	212

## Context and Documents

A Bibliometric-Based Semi-automatic Approach to Identification of Candidate Thesaurus Terms: Parsing and Filtering of Noun Phrases from Citation Contexts	
<i>Jesper W. Schneider, Pia Borlund</i> .....	226

Context Matters: An Analysis of Assessments of XML Documents	
<i>Nils Pharo, Ragnar Nordlie</i> .....	238

## Workshops

Developing a Metadata Lifecycle Model	
<i>Jane Barton, R. John Robertson</i> .....	249

Evaluating User Studies in Information Access <i>Alex Bailey, Ian Ruthven, Leif Azzopardi</i> .....	251
<b>Author Index</b> .....	253

# Wittgenstein, Language and Information: "Back to the Rough Ground!"

David C. Blair

University of Michigan Business School

## 1 Why Language? Why Philosophy? Why Wittgentsein?

First of all, why are the issues of language and meaning important to the study of information systems? Information systems are, of course, tools that are used to search for information of various kinds: data, text, images, etc. Information searches themselves inevitably require the searcher to ask for or describe the information he or she wants and to match those descriptions with the descriptions of the information that is available: in short, when we ask for or describe information we must *mean* something by these statements. This places the requests for information as properly within the study of language and meaning. Surely, requests for information, or descriptions of available information, can be clear or ambiguous, precise or imprecise, just as statements in natural language can. In short, understanding how requests for, and descriptions of, information work, and, more importantly, how they can go wrong, is an issue of language, meaning and understanding.

Why, then, is the focus of this discussion on philosophy? I'm turning to philosophy of language for the principal reason that its *main* concern is with how we *mean* what we say—how language actually works? Another reason why the philosophy of language is particularly pertinent for the present discussion is that for philosophy in general, and Wittgenstein in particular, there is no sharp boundary between understanding language and cognition—how we understand language is closely coupled with how we understand things in general. Not only language, but understanding is important for information systems, too, since information systems are often used to help us understand things better. Since the approach of philosophy of language is the fundamental examination of the issues of meaning, if there are any clear insights into our understanding of meaning, they will likely be found here first. This is why the philosophy of language is so important to the investigation of information retrieval systems.

Why is the philosophy of Wittgenstein particularly important for the study of information retrieval systems? That is, why not just survey the pertinent sections of the Philosophy of Language in general? There are many philosophers of language, and many philosophical theories which have contributed to our understanding of meaning in language. Why should we concentrate our efforts on Wittgenstein's, admittedly difficult, philosophy of language? Surely there are other, easier, routes to furthering our understanding of language and meaning.

But Wittgenstein is unique among philosophers in the following respect: early in his career he was the consummate logician, the intellectual heir apparent to the pioneering logical work of Gottlob Frege and Bertrand Russell. Frege and Russell believed that

ordinary language was not precise enough to represent the complexity and subtleties of meaning that were becoming increasingly important for analytic philosophy. Russell believed that the goal of analytic philosophy was to clarify what we say about the world. Analytic philosophy should take its inspiration from what Russell believed was the rigor of the scientific method. Since different branches of science often needed their own representational systems to express factual scientific relationships clearly, philosophy would need a similar rigorous representational system to make what it could assert perfectly clear, or so Russell & Frege thought.

What we needed, they believed, was a logical language that could faithfully model these complexities and subtleties of expression, and could be used to clarify whether statements of fact were true or false—a language that could be used to bring out and make explicit the underlying logic of language. Early in his career, Wittgenstein was sympathetic with this view of language, believing, like Russell and Frege, that language could be made more precise through the use of formal logic. In his introduction to Wittgenstein's first published work, *Tractatus Logico-Philosophicus*, Russell describes Wittgenstein as being “concerned with the conditions which would have to be fulfilled by a logically perfect language”. Russell goes on to describe a logically perfect language as one which “has rules of syntax which prevent nonsense, and has single symbols which always have a definite and unique meaning”.

But as Wittgenstein's thought matured, he began to have serious misgivings about the ability of logic to model or represent the complex and subtle statements of language. Not only was logic inadequate to this task, he thought, ordinary language itself was, if used properly, the best possible medium for linguistic expression, philosophical or otherwise. In short, Wittgenstein's thought evolved from a belief that problems of meaning in language could be clarified by logically analytical methods to a realization that many of the unclaritys of language were a result of removing statements from the context, practices and circumstances in which they were commonly used—what Wittgenstein called our “Forms of Life.”

What determined the truth or meaning of a statement was not some underlying logic, but how the statement was used and what circumstances it was used in. Ambiguities in language are clarified, not by logical analysis, but by looking at how the words or phrases in question are used in our daily activities and practices. Wittgenstein's transition in his view of language is important for the study of information systems for the following reason: our current most widespread model of information systems is the computer model, in particular, the “data model” of information. This has been a very successful and robust model that has had a remarkably long history of implementation. Computers are, in a fundamental sense, logical machines, so we might say that the current most popular model for information systems is the *logical model*. This logical model, as we will show, has worked well for providing access to the precise, highly determinate content of our data bases—things like names, addresses, phone numbers, account balances, etc. But as more and more of our information is becoming managed by computerized systems we find that we must provide access to less determinate information, like the “intellectual content” of written text, images, and audio recordings—for example, searching for information that analyzes the economic prospects of Central European countries, or information that evaluates the impact of government regulation on

small businesses. These kinds of access are not as well served by the logical data model of information, as one can easily see when trying to find some specific subject matter (intellectual content) on the World Wide Web using an Internet search engine.

Current information systems are in some way, the victims of the success of the more determinate data model of information. The logical/data model of information has become the Procrustean Bed to which many information systems are forced to fit. The effort to fit language and information to the logical model was justified because it was assumed that, as Russell and the early Wittgenstein believed, there is an underlying logic of language that governed its correct usage—an underlying logic which must be uncovered if we wanted to insure the clarity of expression. On this view, information systems used to provide access to "intellectual content" are just sloppy or imprecise versions of data retrieval systems. But it was one of Wittgenstein's clearest reassessments of his early philosophy when he said that "...the crystalline purity of logic was, of course, not a result of investigation; it was a requirement"—that is, the logic that Russell and Frege sought to uncover in their analysis of language, did *not* exist latently in language waiting to be uncovered.

The logic of language was something that was a requirement for the analysis *to begin with*—it was something that was imposed on language. Just as Wittgenstein began to have misgivings about the applicability of the logical model, with its requirement for the strict determinacy of sense, to all aspects of language and meaning, some are now having misgivings about how applicable the logical/data model of information is to the more complex and subtle problems of access to less determinate information such as the "intellectual content" of written text, images and audio recordings, a kind of access becoming increasingly widespread as more and more of our information starts out in machine readable form. For the data/logical model to be applicable to all information systems, it is *required* that the information on the system be represented in extremely precise or determinate ways. But this process will have the effect, not of making better, "more precise" information systems, but, in the case of the search for "intellectual content," of making dysfunctional information systems—systems which are insensitive to the subtleties of language that are required for highly specific access to intellectual content, especially on large systems. As long as we believe that the precision of representation for data retrieval is possible *for all information systems*, we will run the risk of building such dysfunctional systems.

## 2 Surveying Wittgenstein's Landscape

"...we don't start from certain words, but from certain occasions or activities." [LC p.3]

"Let the use of words teach you their meaning." [PI p.220]

**"If a lion could talk, we could not understand him."** [PI p. 223]

"The best example of an expression with a very specific meaning is a passage in a play."

"When I think in language, there aren't "meanings" going through my mind in addition to the verbal expressions: the language is itself the vehicle of thought." [PI § 329]

"Our language can be seen as an ancient city: a maze of little streets and squares, of old and new houses, and of houses with additions from various periods; and this

surrounded by a multitude of new boroughs with straight, regular streets and uniform houses.” [PI § 18]

“Many words...then don’t have a strict meaning. But this is not a defect. To think it is would be like saying that the light of my reading lamp is no real light at all because it has no sharp boundary.” [BB p. 27]

“Frege compares a concept to an area and says that an area with vague boundaries cannot be called an area at all. This presumably means that we cannot do anything with it.—But is it senseless to say: “Stand roughly there”? [PI § 71]

“If a pattern of life is the basis for the use of a word then the word must contain some amount of indefiniteness. The pattern of life, after all, is not one of exact regularity.” [LWPP I § 211]

“We want to establish an order in our knowledge of the use of language: an order with a particular end in view; one of many possible orders; not the order.” [PI §132]

“My method is not to sunder the hard from the soft, but to see the hardness of the soft.” [NB p.44]

“The more narrowly we examine language, the sharper becomes the conflict between it and our requirement. (For the crystalline purity of logic was, of course, not a result of investigation; it was a requirement.) The conflict becomes intolerable; the requirement is now in danger of becoming empty.—We have got onto slippery ice where there is no friction and so in a certain sense the conditions are ideal, but also, just because of that, we are unable to walk. We want to walk; so we need friction. Back to the rough ground!” [PI § 107]

### 3 Wittgenstein’s Main Views of Language

“Meanings” are not linked to words.

“Meanings” are not concepts or any other single thing.”

To understand the meaning of a word is not to have some definition in your head, but to be able to use the word correctly in the activities and practices in which it is normally used.

The context of usage is essential for understanding language.

Indeterminacy in language is inevitable, but is not the result of sloppy or irrational usage.

### 4 Types of Indeterminacy in Information / Content Retrieval:

- Semantic Ambiguity
- Category Overload
- Language Productivity

# Text, Co-text, Context and the Documentary Continuum

Elisabeth Davenport

School of Computing Napier University,  
Edinburgh EH10 5DT  
e.davenport@napier.ac.uk

**Abstract.** The paper is concerned with ways in which we understand context. In mainstream LIS, context is construed as environment or situation, a place where work gets done, supported more or less by information objects that are retrieved from a different space. The resulting separation of object and agent underlies two significant lines of work in the LIS domain: the search for optimal access to objects and the description of human information behaviour. Performance measurement dominates the former; the latter has led to elaborate and universalist models that have little discriminatory power and whose validity is difficult to establish. Both groups are pre-occupied, in their own way, with matching agent and object, or with relevance, though the question of 'relevant to what?' has many different answers - tasks, life mastery, leisure interests and so on. A recent 'call to order' here suggests that 'tasks and technology' should be the focus of LIS efforts, as these can at least support the validation of empirical work.

Context, however, may be understood differently, in terms of the texts that surround or are linked to a specific text that demands attention at a given moment. (For the purposes of this argument, context is not commensurate with hypertext; hypertext is a technology that supports the recording of context). A text (in terms of functional linguistics) is a meaningful unit of language, and texts vary in size - a clause, an article, a recipe, and (stretching the concept for the purposes of the argument presented here) a collaboratory. To focus thus on texts is not to take the side of 'objects' versus 'agents' - the agent is present in the text and achieves certain ends through texts. These are accomplished within a system of encoding (choices about what is appropriate) that makes meaning possible among those who share the code. For those in the know (members of a group, a discipline, a community) encoding is implicit. Outsiders, in contrast, need to work hard to grasp why choices have been made - what is linked with what, what refers to what, in what ways cohesion is achieved, and this work is often described in terms of literacy. Sociolinguists use a number of terms in discussing these issues - co-text (coding drives textual links to text) and social context (coding drives actions that are non-textual). These are dimensions (along with content) of a continuum - text and function, text and use are tightly coupled.

Within LIS, there are a number of approaches to clustering documents that are based on inter- and intra-textual analysis; this 'co-textual' work, however, has not been perceived as dealing with 'context'. Citation analysis is the most salient of these, with the journal article as the standard unit. It makes visible patterns of social choice and aggregation, and reveals the social encodings that characterise different domains. Citation analysis is concerned with use: sequences of 'uses' establish the texture of a domain and these threads can be traced at a very fine level of detail - a clause (the smallest unit of text) for example, may be tracked across a sequence of documents. Citation analysis may be seen as a prototype



for recent systems for social filtering and personalisation (recommender systems, reputational systems). These systems change the texture of work, by providing shifting sets of possibilities for action.

So why foreground an existing technique? The last part of the paper suggests that co-textual analysis is particularly timely in a world of large scale digital collections (or archives) - the aggregates associated with cyber-infrastructures and grid technologies, with e-science, e-learning, e-government, e-commerce. These collaborative spaces for project or service work are constructed on a scale that complicates understanding. They are also complex texts, whose emergent effects contribute to problem framing and problem solving. Existing tools for 'knowledge discovery' in such environments focus on the management of content, or objects that are described and classified (meta-data and meta-languages). But large collaborative texts also require tools for explication (based on co-textual analysis), that address transitivity and trace threads over time, mapping, for example, shifts in the argument of a complex text or shifts in the structure of a complicated document over versions. Visualisations of this kind can be used to track specific couplings of text, co-text and context at different levels of aggregation (what has recently been described as 'textography'); they are also important drivers of literacy - at the level of the domain, or the project.