Better Search Applications Through Domain Specific Context Descriptions: a Position Paper

Corrado Boscarino
Centrum Wiskunde & Informatica
Science Park, 123
1098 XG Amsterdam, The Netherlands
corrado@cwi.nl

Arjen P. de Vries*
Centrum Wiskunde & Informatica
Science Park, 123
1098 XG Amsterdam, The Netherlands
arjen@acm.org

Jacco van Ossenbruggen†
Centrum Wiskunde & Informatica
Science Park, 123
1098 XG Amsterdam, The Netherlands
jrvosse@cwi.nl

ABSTRACT

There is a wide agreement that a user centred approach to IR applications design outperforms system centred ones. A classic understanding of this design approach, and specifically of its underlying notion of context, appeared however insufficient to explain the results of a pilot experiment. We recognise the importance of context, but we define context differently by means of a domain theory: a conceptualisation of the domain at hand, preferably developed within the same community which users belong to.

1. INTRODUCTION

We believe that there is a need to escape the narrowness of existing system centred application development and evaluation frameworks, towards designs that are more open and alert to the human user, whose purposes they are supposed to serve. Karen Spärk Jones [10] raised the same issue already in the late Eighties, arguing in favour of lines of research in Information Retrieval (IR) that go beyond probabilistic weighting and of evaluation paradigms that step outside the entrenchment of the laboratory. These are still present day challenges for IR research [4]: IR applications should address a more realistic environment comprising real users with all their expectations, naiveties, affections, clear-cut ideas or just loose or even incoherent judgements.

In IR the most influential and successful attempts to take a user centred stance and to propose an alternative to a classic document space model [14] have been developed from a cognitive perspective [3]. We regard an IR system as supporting a communicative act between an expert on a certain matter, who made information available about her knowledge of the world, and a system's end-user, who at a certain moment became aware of an anomaly in her epistemic state. When a user issues a query onto the system in an effort to eliminate that anomaly [2], we may interpret IR in terms of linguistic interchange. Language is a paradigmatic form of human interaction and several authors describe a cognitive view, "a forceful theoretical foundation for IR interaction and human-computer interaction (HCI) in general" [8] in terms of linguistic notions.

A cognitive view amounts then to recognising that, while a sentence, an image or more generally a *sign* may have many different meanings, interpretation relies on a cognitive process upon receipt of this sign, which is only potentially informative [8]. The goal of search applications upon this view is "to give access to [these] plausible means or values as, given the situation, may entail transformations of a cognitive state, thus providing information in a pragmatic sense via context" [8].

A cognitive view of IR extends therefore the communication system setup of [2], which is inspired by quantitative models of information exchange [16] such as those for transmissions of electrical signals over noisy channels, into a more complex setting where *sender* and *receiver*, or *speaker* and *hearer* in a classical model of linguistic exchange, are turned down in favour of agents, who engage in an interactive game. The focus of this 'linguistic turn' in IR, a shift from considering documents as labels to regarding them to be part of a sense making process is an analysis of a linguistic and behavioural game: a pragmatics of meaning, that is a model of the way signs are used during information exchange. A cognitive approach explicitly takes an internal perspective [1] by addressing the mental states of agents who play the game, with its set of mutually agreed upon rules, strategies that each agent is determined to employ and criteria of assessing a positive conclusion of the game.

Many approaches to design often derive a potential to gather consensus around their underlying notions from common sense intuitions: obvious properties of reality with which everyone is expected to agree. These unquestioned postulates may not fully endure a more thorough analysis and can be challenged by empirical evidence when these general theoretical frameworks are translated into practice. A tacit assumption in a pragmatics of IR is that understanding the process through which a world view mediates interpreta-

*Also affiliated to the Delft University of Technology
†Also affiliated to the VU University Amsterdam

HCIR 2010 New Brunswick USA

tion amounts to understanding the characteristics of a world where an act of interpretation takes place.

This short position paper reports on how, during the development of a pilot user study designed to improve our knowledge on combining information from different sources, we came to evince a lack of coherence at the very bottom of IR systems design upon a cognitive view. Exchanging a hypothetical external observer monitoring communication for an analysis of a subjective mental representation of how a game of exchange is being played leaves a scientist seeking to develop cognitive models of information exchange in an ambiguous position: those internal mental models are still supposed to be inferred from observed behaviour. A context, which linguistic indexicality refers to, becomes then a comprehensive set of material, social, cultural and psychological characteristics, which deterministically settle observed behaviour.

We believe that this rather implicit assumption of considering mental representations to be observables within a cognitive model is the main motivation of most efforts in developing a wide range of techniques to improve our knowledge of a user's world. Think aloud protocols, ethnographic studies or extrapolations of simulated work tasks [9]: they all aim to carefully charting a user's environment. Provided that we take a sufficiently unbiased view, empirical data are the means to understand user's behaviour. Conversely, our inability to make sense of some experimental result or to deduce a correct set of design requirements for a new application is a consequence of a lack of empirical material or a flaw in our methodological tools. Such ground assumptions, given the weight of these studies in modern IR research, should be investigated.

We are confident that social structures do matter, but we argue that a linguistic interpretation of the cognitive view in IR seems to point against their observability from user performance. While the discourse analysis community recently suggested that "*there is no direct link between situational or social structures and discourse structures*" [5], no comparable paradigm shift has yet emerged in accounts of IR as discursive practice.

IR is a highly technologically advanced linguistic game and we must grant the participants to this game more skills than current accounts do: apart from intuitively grasping the material and social background of linguistic interaction, players are proficient in abstracting on a set of rules by following which a game can be successfully completed. We will suggest to investigate one of these abstract and domain specific models of a particular discourse that happens to underlie our prospective application as the key of a solution to correct interpret the results of our user study.

The following section explains how we designed and performed our pilot user study and how it challenged our views on what should be regarded as context of the linguistic IR performance. Given the limited space that we have, we will not give much details and we will present the experiment as an empirical trigger for a more abstract claim. Section 3 introduces our solution and in the last section we propose some directions to generalise this conclusion to other IR applications. Our central claim is that we should consider, not only analyses of a user's environment as it appears to a computer scientist approaching a new design, but also the different understandings of a domain of interest that were developed in the same community which a user belongs to.

2. STAGING A LINGUISTIC IR GAME

Acknowledging a view that regards HCI and IR as technology enabled linguistic exchanges and at the same time one that considers a cognitive process to provide the main contribution to sense making, amounts to stress the functional role of language, that is how language is used during interactions, rather than its capacity as information carrier: "communication should be grounded in the pragmatic practices by which speakers use language" [12]. Understanding an IR speech act requires to investigate the situation in which this act takes place. Whitin this theoretical framework we developed a user study aiming to understand how to support combined search for documents which contain information from two repositories: one of textual documents and another of images.

Our ground hypothesis was that, because of what Ingwersen termed 'cognitive free fall' [8], syntactic and semantic components of meaning captured by classic text retrieval language models would trigger characteristic pragmatic features of discourse: "*intentionality*, meaning, implicit *context* and potential *informativeness* underlying the generated and communicated message [that] are *immediately lost* [...] have to be rebuilt and recovered" [8]. In order to investigate this reconstruction process, we designed two different algorithms to search onto textual documents contained in the Keesings[1] archive and pictures available in the Spaarnestad[2] collection. The first repository is a Dutch archive of historical chronicles and the second one contains for the largest part press pictures annotated with free text and some labelled entities such as persons and publication date in one XML file. In both algorithms we used the free text annotations to match their correspondent pictures to one textual document. In a first algorithm the PF/Tijah[3] text retrieval software provided a relevance ranking of the annotations to a textual document: the pictures corresponding to the top 5 ranked annotation, provided that they were above a 0.5 threshold, were chosen for the user study. This is a measure of how much a set of pictures are about a certain textual document.

A second, two-step algorithm aimed to incorporate in the 'aboutness' ranking the information available in the textual collection about a picture. First we generated what can be thought of as a picture's representation from the point of view of an external observer, who has complete knowledge of the textual collection. For every picture we extended its annotation with a set of paragraphs from the text collection ranked top 10 by the PF/Tijah software, again provided that they were above a 0.5 threshold, without regard of which particular document they belong to.

A second step is similar to the first algorithm: pictures associated with the top 5 ranked annotations against the complete document collection are considered to match the information contained in a textual document. We invited 4 people, in different extents familiar with the text collection, to participate in a user study. We asked them to arrange a small number of textual documents and pictures, each deemed relevant to the textual document by at least one algorithm, in joint excerpts containing one textual document and at least more than one picture. They have been left free to search with a keyword based search engine onto the

[1] `http://www.kha.nl`
[2] `http://www.spaarnestadphoto.nl`
[3] `http://dbappl.cs.utwente.nl/pftijah`

image repository and compare the quality of the images provided by our two algorithms, discussing the process through which the joint documents were being generated. Because of our assumption on the existence of syntactic and semantic constraints to interpretation, we expected that the two algorithms would trigger different behaviours when employed in a practical search scenario.

As the information in the archive represents a collective effort of many different people, who mediated between issues as different as aesthetic aspects, scientific rigour, economic feasibility or maintenance requirements, we expected that, on average over the different topics and despite of users' different personal backgrounds, one journalist, a historian, a photographer and a member of the Keesings administrative staff, a clear preference for the second algorithm in all sessions would emerge; only this second algorithm was, at least in principle, designed to grasp some contextual features.

3. THE PROBLEM

Users' performances were audio-recorded and analysed. Since methods of investigation depend on prior methodological stances [17] and results on our framing of the user, the importance of making methodological choices accountable should not be underestimated: we selected *practitioner's profiles* [7] as a tool to achieve our goals of putting practical discursive performance centre stage. Interviews with the participants to the user study concentrate on *how* an action is performed rather than on *why*: instead of asking "why are you doing this or that", the interviewer focused on asking "how do you achieve this or that result".

We found that many of these practice narratives exhibited what we termed signposts of experiences: expressions that denote familiarity with a certain topic. "Nobody knows who is he", says one user referring to Brian Epstein when commenting on using a certain picture to illustrate a text about the Beatles, "I believe you want to show that: I can still remember to have seen him cycling" says another user about a very famous Dutch trainer and a not-so-famous professional cyclist himself. Every time we report an occurrence of a signpost we assign a label to the correspondent user: we call 'expert' on a matter a user who speaks out a signpost in agreement with the information that is present in the textual collection; a case of disagreement is labelled as 'non-expert'. It is important to remember that, during analysis these labels simply denote a discursive pattern and the process of inferring users' mental models builds upon it, but it is by no means identified with it. We draw from the same methodological viewpoint, which has already demonstrated its utility in HCI [13], promoting a phenomenological account of social constructions of reality: 'expert' and 'non-expert' are categories in phenomenological sense [11], both a regular pattern of reality and a mental classification at the researcher's side. Actions and discursive evidence were thus mapped to these two categories. We found that 'experts' were showing a preference for the two-steps algorithm while, against also our expectations, 'non-experts' were indifferent to the two algorithms.

Making sense of these tracings of discourse analysis and of the categorisation induced by the two algorithms turned out to be impossible. Data are available to show that users should react differently to the two algorithms, because of the way the algorithms are constructed. Data are also available to show that the categorisation of 'expert' and 'non-expert' corresponds to measured patterns: one can criticise the terms, but must agree on the content, even by assigning more neutral labels such as 'a' and 'b'. We are puzzled by an apparent paradox: we can argue separately in favour of two categorisations, two divisions of the information interaction space, but we cannot explain their combined effect.

Our first reaction was to question the scope of this experiment. Gathering more data or improving our investigation tools leads, so goes the received view, to a secure determination of all the relevant parameters of the cognitive process underlying interpretation. In our case either the algorithms were not grasping any interesting contextual feature, or the collection did not represent a shared knowledge among all our users, or our panel was too limited, or our methods too superficial or some even more dangerous combination of all these factors flawed our judgement. While we recognise that all these issues may have conspired against us and we agree that much more extended experiments may well overturn our results, we suggest that another possibility has been for too long, too easily discounted by information interaction researchers: the possibility that "discourse and actions are not immediately observable at all" [5], that analysis is mediated by a theoretical model and that the validity of a design is only relative to that theoretical model.

4. THE SOLUTION

In this last section we show a positive example in favour of the claim that we can make sense of our results by limiting the scope of our analysis rather than widening it. We take then a step back, recasting this problem in linguistic and cognitive terms in order to find out where we made an unwarranted assumption that generated a paradox.

What we did not justify at any point is why an analysis of the setting of a situation allows a complete understanding of the cognitive process, hence of interpretation in IR: "it is a widespread misconception [...] that social situations and their properties [...] exercise *direct* and unmediated influence on language use" [5]. We can paraphrase this critique: even if improve our investigation techniques, for example exchanging practitioner's profiles for ethnographic or log-data analysis, there is no guarantee that we understand what a user is doing.

Once we recognise this issue, what seemed a puzzle of interpretation of empirical data, becomes a more fundamental and striking impasse. On one side we cannot predict any particular matching from a knowledge, how accurate it may be, of user performance. On the other hand the matchings that are relevant for a task are shared at least within a small community: there is no point in over fitting any algorithm to comply with the requirements of one unique user. Notice that we do not need a cognitive framework to show this particular point: even in a noisy channel account of IR, or for that matters any communication channel, even the most abstract [15], perfect replication of information at the receiver side, that is in IR perfect replication of the information possessed by a knower at the seeker's side, is theoretically untenable: it implies identifying a knower with an information seeker and therefore allowing in our model only users with a perfect knowledge of the system.

There is no hope to resolve this impasse within a cognitive and linguistic model. The subjective construction of interpretation in IR that is modelled as a chain of cognitive state changes during interaction [8] irremediably conflicts with a

requirement of any linguistic pragmatics: a totally private use of language must be denied. At a certain point the chain of pragmatic explanations must end [18].

Our suggestion to solve this problem is to adapt an insight from context theory: there are mental interfaces that mediate discursive performance and "what is observably done or said is only the tip of a iceberg of a communicative event" [5]. We also suggest that characterisations of these mental interfaces or context models should be sought within the same community, which users belong to; most often outside computer science. The results we may achieve as computer scientists are therefore sound only with respect to that characterisation.

In our case *remediation* theory, an account of the process "by which new media refashion prior media forms" [6], provides a local solution to the paradox by adding one additional constraint: remediation depends on the existence of a *prior* personal agenda. Those who do not have any background knowledge on the matter at hand, cannot have a personal agenda: interpretation is then based on the information provided by the system at the time the interpretation act is performed. While we witnessed the act of joining different media, only one group, that labelled as 'expert', was performing an act of remediation, because only one group complied with this additional requirement. If we limit the scope of our algorithms to the case of remediation, we then may conclude that the second algorithm performed better. We are not caught into the impasse because we have to accept it as inherent to the process that we want to support as the 'double logic of remediation' [6].

5. CONCLUSIONS

The initial aim of our user study was to develop an application to allow end-users to retrieve joint documents comprising information from two different repositories, one of historical documents and one photo archive. While the cognitive viewpoint in its linguistic interpretation is confirmed as a very powerful approach to design, the key to solve the issues raised by this task is to modify the received notion of context of interpretation: we derive a sharp distinction between social properties of a situation and cognitive processes that unfold during interaction. Interpretation is still "*mediated* by a system of categories or concepts which, for the information-processing device, are a *model* of his *world*" [3, citing De Mey], but a world description, however accurate it may be, is not enough to completely understand this process of mediation. A subjective construction determines the outcome of this communication act.

Working out the consequences of this paradigm shift results in suggesting an alternative design flow for IR applications. Once we abandon a positivistic belief in the direct observability of discourse and action we need other grounds for motivating applications. Once the ideal of investigation methods, which at least in the limit can converge to general solutions is recognised as illusory, researchers are left with a weaker, but more reliable notion: that of a world model as conceptualised by a theory specific to the particular domain which an end-user belongs to. Design should then start with a first, critical step: a choice to employ a domain theory for the particular scenario rooted in the research tradition and in the practice of the domain at hand and engaging its domain specific conceptualisation and even its proper jargon.

6. REFERENCES

[1] B. G. Bara and J. Douthwaite. *Cognitive Pragmatics: The Mental Processes of Communication*. MIT Press, 2010.

[2] N. J. Belkin. Anomalous states of knowledge as a basis for information retrieval/ Etats anomaux des connaissances comme base du repérage de l'information. *The Canadian Journal of Information Science*, 5:133–143, 1980.

[3] N. J. Belkin. The cognitive viewpoint in information science. *J. Inf. Sci.*, 16(1):11–15, 1990.

[4] N. J. Belkin. Some(what) grand challenges for information retrieval. *SIGIR Forum*, 42(1):47–54, 2008.

[5] T. A. v. Dijk. *Society and Discourse: How Social Contexts Influence Text and Talk*. University press, Cambridge, 2009.

[6] A. Erll and A. Rigney. *Mediation, Remediation, and the Dynamics of Cultural Memory*. Walter de Gruyter, 1st edition, 2009.

[7] J. Forester. Exploring urban practice in a democratising society: opportunities, techniques and challenges. *Development of southern Africa*, 23(5), dec 2006.

[8] P. Ingwersen. Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation*, 52:3–50, 1996.

[9] P. Ingwersen and K. Järvelin. *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer-Verlag, Secaucus, NJ, USA, 2005.

[10] K. S. Jones. A look back and a look forward. In *SIGIR '88: Proceedings of the 11th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 13–29, New York, NY, USA, 1988. ACM.

[11] K. Kwang-ki and T. Berard. Typification in society and social science: The continuing relevance of Schutz's social phenomenology. *Human Studies*, 32(3):263–289, 2009.

[12] K. B. Ng. The applicability of universal pragmatics in information retrieval interaction: a pilot study. *Inf. Process. Manage.*, 38(2):237–248, 2002.

[13] K. B. Ng. Toward a theoretical framework for understanding the relationship between situated action and planned action models of behavior in information retrieval contexts: contributions from phenomenology. *Inf. Process. Manage.*, 38(5):613–626, 2002.

[14] G. Salton. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley, 1989.

[15] M. Serres. *Le parasite*. John Hopkins University, London, 1982.

[16] C. Shannon, N. Petigara, and S. Seshasai. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

[17] T. D. Wilson. Alfred Schutz, phenomenology and research methodology for information behaviour research. In *ISIC 2002*, pages 193–207, Lisbon, Portugal, 2002.

[18] L. Wittgenstein. *Philosophische Untersuchungen - Auf Grundlage der Kritisch-genetischen Edition*, volume 1372. Suhrkamp, Frankfurt a. M., September 1953.