Fundamental Properties of Aboutness

Peter Bruza
Distributed Systems Technology Center
Building 78, Staff House Road
University of Queensland, Qld 4072
Australia
bruza@dstc.edu.au

Dawei Song
Department of systems Engineering and Engineering Management
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong
dwsong@se.cuhk.edu.hk

Kam-Fai Wong
Department of systems Engineering and Engineering Management
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong
kfwong@se.cuhk.edu.hk

1. INTRODUCTION

Information retrieval (IR) is a reasoning process which is assumed to be driven by determining aboutness (|=) between two information carriers (i.e., document and query). Thus, the study of aboutness will be very helpful to set up the theoretical foundations of IR. Aboutness is modeled as a binary relation over the information carriers (IC). Early studies viewed aboutness as a form of entailment. We regard aboutness as a broader notion. Recent attempts have been made to formalize properties of aboutness which can be expressed as postulates (rules) in terms of information containment, composition and preclusion. However there is yet no consensus regarding this framework except that it should be logic-based [7]. Although a number of aboutness properties are commonly discussed in the literature, e.g., reflexivity, transitivity, symmetry and left (right) monotonicity, etc., there is thus far no agreement on a core set of aboutness postulates, e.g., Hunter deems aboutness to be irreflexive [6] whereas Huibers deems it reflexive [5]. The disagreement stems partially from the framework chosen to formalize aboutness. Hunter uses default logic whereas Huibers uses situation theory. Once the framework has been fixed, certain aboutness properties are implied by it. Moreover, some properties, e.g., transitivity and symmetry, etc., may be sound only within certain IR models, and some of them may lead to negative effects to the effectiveness of IR system. In this article, by adopting a very simple framework, we attempt to gain enough freedom to propose and discuss a wide range of aboutness postulates without being bound too much by the underlying framework. Cleverden cites experiments wherein the agreement between subjects judging documents with respect to a query was around sixty percent [4]. This suggests that aboutness have a subjective component. However, there also seems to be a core of agreement, which, in our opinion is amenable to formal treatment. Thus, the purpose of this article is to consider aboutness from a fundamental, commonsense perspective, to shed light on the nature of aboutness by formalizing properties describing it, and to define a set of reasonable (hopefully sound) properties of aboutness, which is independent of any given IR model.

2. PRELIMINARY FRAMEWORK

Our framework is defined as {IC, $\rightarrow$, $\oplus$, $\perp$}, with the following properties:

(1) Reflexivity: A$\rightarrow$A
(2) Transitivity: A$\rightarrow$B, B$\rightarrow$C $\Rightarrow$ A$\rightarrow$C
(3) Asymmetry: A$\rightarrow$B doesn't imply B$\rightarrow$A
(4) Containment-Composition (CC): A$\oplus$B$\rightarrow$A; A$\oplus$B$\rightarrow$B
(5) Absorption: A$\rightarrow$B $\Rightarrow$ A$\oplus$B=A
(6) Non-conflict containment (NCC): A$\rightarrow$B $\Rightarrow$ A $\not\perp$ B
(7) Containment-Preclusion (CP): A$\rightarrow$B, B$\perp$C $\Rightarrow$ A$\perp$C

Where A, B, C $\in$ IC; information containment (A$\rightarrow$B) models the information is explicitly and implicitly nested; information composition (A$\oplus$B) models A and B can be composed to more complex information carrier; information preclusion (A$\perp$B) means A clashes, or contradicts, with B.

3. "AXIOMATIZING" ABOUTNESS

The following are argued as commonsense aboutness properties:

(R) *Reflexivity*

(AS) *Asymmetry*

(AC) *Aboutness Consistency*:

$$\frac{A \models B}{A \perp B}$$

(C) *Containment*:

$$\frac{A \xrightarrow{X} B}{A \models B}$$

where X is the maximal steps of transitivity of information containment to keep aboutness relation. Brooks found that it is approximately two-step [1].

(CT) *Cut*:

$$\frac{A \oplus B \models C \quad A \models B}{A \models C}$$

(M) *Mix*: $\dfrac{A \models C \quad B \models C}{A \oplus B \models C}$

(A) *And:* $\dfrac{A \models B \quad A \models C}{A \models B \oplus C}$

(QLM) *Qualified Left Compositional Monotonicity*:

$$\frac{A \models B \quad B \perp C \quad C \not\rightarrow A}{A \oplus C \models B}$$

(QRM) *Qualified Right Compositional Monotonicity:*

$$\frac{A \models B \quad A \perp C \quad B \not\rightarrow C}{A \models B \oplus C}$$

(E) *Equivalence:* $\dfrac{A \models B \quad B \models A \quad A \models C}{B \models C}$

4. NON-ABOUTNESS

Several authors have investigated this notion [2, 3, 5, 6]. Information filtering is an example of a situation where reasoning about the non-aboutness of incoming documents with respect to the user profile may be easier than reasoning about their aboutness. We drop the closed world assumption regarding $\models$ and determine non-aboutness ($\not\models$) via constructive means. In the following we describe commonsense properties of non-aboutness:

(P) Preclusion: $\dfrac{A \perp B}{A \not\models B}$

(N-C) Containment Non-aboutness: $\dfrac{A \xrightarrow{>x} B}{A \not\models B}$

(P-NA) Preclusion Non-aboutness:

$$\frac{A \models B \quad B \perp C}{A \not\models C}$$

(S-NA) *Symmetry Non-aboutness*

Proposition 1 *P-NA and S-NA are derivable properties.*

5. INTERACTION BETWEEN ABOUTNESS AND NON-ABOUTNESS

The following properties are normative rules depending on whether an optimistic or pessimistic stance is adopted. It is assumed that one is either an optimist or a pessimist.

(OL) *Optimistic left:* $\dfrac{A \not\models B \quad C \models B \quad A \not\to C}{A \oplus C \models B}$

(OR) *Optimistic right:* $\dfrac{A \not\models B \quad A \models C \quad B \not\to C}{A \models B \oplus C}$

(PL) *Pessimistic left:* $\dfrac{A \not\models B}{A \oplus C \not\models B}$

(PR) *Pessimistic right:* $\dfrac{A \not\models B}{A \not\models B \oplus C}$

6. COMPLETENESS, CONSISTENCY AND SOUNDNESS

The completeness of an aboutness reasoning system means for any two arbitrary information carriers A and B, the system must be able to conclude either $A \models B$ or $A \not\models B$.

Proposition 2 *The aboutness and non-aboutness system {R, C, QLM, QRM, E, P, N-C} is complete.*

It would be undesirable for an aboutness inference system to be inconsistent, i.e., $A \models B$ and $A \not\models B$ cannot be true at the same time.

Hypothesis[1] *The commonsense aboutness and non-aboutness systems together with a pessimistic stance {PL, PR} are consistent*

Verifying soundness cannot be approached as is traditionally done in logic: Aboutness is a fuzzier notion than truth. Moreover, unsoundness may be tolerated in order to promote recall of an IR system (i.e., via optimism). The degree of unsoundness may turn out to be a more pertinent question than whether an aboutness system is sound or not.

[1] We use "hypotheses" here because at the moment this question is still being worked on.

7. INTENDED APPLICATIONS

Aboutness is an important area in theoretical study of IR. Our belief is that a better understanding of aboutness will lead to significant breakthrough in IR theory and more effective IR systems. Moreover, it could be applied to the following fields:

- *IR functional benchmarking*. The traditional empirical methods (performance benchmarking) are good at evaluating the performance of a system, they are unable to assess its underlying functionality. This can be overcome by aboutness based functional benchmarking.
- *Query expansion*. This is also a reasoning process for query-query aboutness decision. The desirable properties of aboutness can serve as a guild to improve the effectiveness of the inference rules within the query expansion process.
- *Intelligent agents*. They can use aboutness, non-aboutness theorems and the interaction between them to help make relevance and non-relevance decision, e.g., in information filtering, the non-relevant documents are first excluded according to the work of non-aboutness agent.

8. FUTURE WORK

IR models often employ various weighting factors. For simplicity we have not considered them. In the future we plan to incorporate them by ordering the initial aboutness relationships and the inferences produced. Among other things, this will allow a more fine-grained analysis of aboutness and non-aboutness. In addition, we will consider "similarity" relation, which can model the document (or term) clustering, and its interaction with aboutness.

9. REFERENCES

[1] Brooks, T.A. People, words, and perceptions: a phenomenological investigation of textuality. Journal of American Society for Information Science. 46 103-115, 1995.

[2] Bruza, P.D. and Huibers, T.W.C. Investigating aboutness axioms using information fields. In Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval. Dublin, Ireland, 112-121, 1994.

[3] Bruza, P.D. and Huibers, T.W.C. A study of aboutness in information retrieval. Artificial Intelligence Review *10*, 1-27.

[4] C.W. Cleverden. The Significance of the cranfield tests on index languages. Proceedings of the 14th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, pp 3-12, 1991.

[5] Huibers, T.W.C. An Axiomatic Theory for Information Retrieval. Ph.D. Thesis, Utrecht University, 1996.

[6] Hunter, A. Intelligent text handling using default logic, In Proceedings of the Eighth IEEE International Conference on Tools with Artificial Intelligence (TAI'96), 34-40.

[7] Lalmas, M. and Bruza, P.D. The use of logic in information retrieval modeling. Knowledge Engineering Review 13 (3):263-295, 1998.

[8] van Rijsbergen, C.J. A non-classical logic for information retrieval. *The Computer Journal 29*, 6, 481-485, 1986.

[9] Sun, R. Robust reasoning: integrating rule-based & similarity-based reasoning. Artificial Intelligence 75 (1995) 241-295.