

A Basis for Information Retrieval in Context

MASSIMO MELUCCI

University of Padua

Information retrieval (IR) models based on vector spaces have been investigated for a long time. Nevertheless, they have recently attracted much research interest. In parallel, context has been rediscovered as a crucial issue in information retrieval. This article presents a principled approach to modeling context and its role in ranking information objects using vector spaces. First, the article outlines how a basis of a vector space naturally represents context, both its properties and factors. Second, a ranking function computes the probability of context in the objects represented in a vector space, namely, the probability that a contextual factor has affected the preparation of an object.

Categories and Subject Descriptors: H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—*Retrieval models*

General Terms: Theory

Additional Key Words and Phrases: Personalization, quantum mechanics, probability, vector-space model

ACM Reference Format:

Melucci, M. 2008. A basis for information retrieval in context. *ACM Trans. Inform. Syst.* 26, 3, Article 14 (June 2008), 41 pages. DOI = 10.1145/1361684.1361687 <http://doi.acm.org/10.1145/1361684.1361687>.

1. INTRODUCTION

An information retrieval (IR) system is concerned with retrieving all and only those documents which are relevant to any information need of any user. IR is intrinsically context dependent; what is relevant to one user in one place at one time can no longer be relevant to another user, in another place, or at another time. In principle, an IR system should be context aware. In practice, classical systems such as search engines are unaware of such a highly dynamic search environment and contextual features are not captured at indexing time, nor are they exploited at retrieval time. Therefore, retrieval may be inaccurate, especially when one-word queries are submitted to the system; they need to

Author's address: M. Melucci, Dipartimento di Ingegneria dell'Informazione, Via Gradenigo, 6-35131-Padova-Italy; email: melo@dei.unipd.it.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701, USA, fax +1 (212) 869-0481, or permissions@acm.org. © 2008 ACM 1046-8188/2008/06-ART14 \$5.00 DOI 10.1145/1361684.1361687 <http://doi.acm.org/10.1145/1361684.1361687>

ACM Transactions on Information Systems, Vol. 26, No. 3, Article 14, Publication Date: June 2008.