

# Unitary Operators on the Document Space

**Eduard Hoenkamp**

Nijmegen Institute for Cognition and Information

hoenkamp@acm.org

Latent Semantic Indexing (LSI) dramatically reduces the dimension of the document space by mapping it into a space spanned by conceptual indices. Empirically, the number of concepts that can represent the documents are far fewer than the great variety of words in the textual representation. Although this almost obviates the problem of lexical matching, the mapping incurs a high computational cost compared to document parsing, indexing, query matching, and updating. This paper shows how LSI is based on a unitary transformation, for which there are computationally more attractive alternatives. This is exemplified by the Haar transform, which is memory efficient, and can be computed in linear to sublinear time. The purported advantages of LSI are thus preserved while the computational costs are drastically reduced.