Information Retrieval(IR) systems are gaining importance due to wide range of applications like recommender systems, search engines, etc., however, most of the IR systems use statistical methods built on top of bag-of-words approach for text retrieval. Graph-of-words approach is an alternative to bag-of-words approach that uses graph theoretic methods to rank keywords and related documents. We represent text documents as graphs whose vertices correspond to the unique terms belonging to the document. The edges represent co-occurrences between the terms. The underlying assumption is that the terms that co-occur have some sort of semantic relationship that can be harnessed for IR systems. The significant terms can be extracted using graph centrality measures. In this book, we have proposed a novel graph-of-words indexing technique using eigenvector scores that uses case separation for Gujarati language. We compared the performance of IR systems of our approach over the classical bag-of-words approach, mean average precision (MAP) values obtained in our experiments show that our approach has shown significant improvement over classical approaches.

Dr. Hardik Joshi is an Assistant Professor with the Dept. of Computer Sc., Gujarat University, India. His research interests include Natural Language Processing and Information Retrieval.

978-620-0-08276-3

Significant Keywords from Gujarati Text

Joshi

Dr. Hardik Joshi

Identification of Significant Keywords from Gujarati Text Documents

LAP LAMBERT Academic Publishing

Dr. Hardik Joshi

Identification of Significant Keywords from Gujarati Text Documents

LAP LAMBERT Academic Publishing

Imprint
Any brand names and product names mentioned in this book are subject to trademark, brand or patent protection and are trademarks or registered trademarks of their respective holders. The use of brand names, product names, common names, trade names, product descriptions etc. even without a particular marking in this work is in no way to be construed to mean that such names may be regarded as unrestricted in respect of trademark and brand protection legislation and could thus be used by anyone.

Cover image: www.ingimage.com

Publisher:
LAP LAMBERT Academic Publishing
is a trademark of
International Book Market Service Ltd., member of OmniScriptum Publishing Group
17 Meldrum Street, Beau Bassin 71504, Mauritius

Printed at: see last page
ISBN: 978-620-0-08276-3

Contents

List of Figures

List of Tables