SEMANTIC BASED PATTERN SEARCH ENGINE

Dr P. AJITHA

Professor/CSE

School of Computing,

Sathyabama Institute of Science and Technology,

Chennai.

Jupiter Publications Consortium
22/102, Second Street, Virugambakkam
Chennai - 600 092. www.jpc.in.net
Email: director@jpc.in.net

Acknowledgments:

Any trademarks, service marks, product names, or named features are assumed to be the property of their respective owners and are used only for reference. There is no implied endorsement if we use one of these terms.

Published by:

Jupiter Publications Consortium, [2024]

SEMANTIC BASED PATTERN SEARCH ENGINE

Dr P. AJITHA

ISBN: 978-93-86388-47-6
First Published: 10th September 2024
DOI: www.doi.org/10.47715/978-93-86388-47-6
Price: 175/-
No. of. Pages: 60
Jupiter Publications Consortium
Chennai, Tamil Nadu, India
E-mail: director@jpc.in.net
Website: www.jpc.in.net

Name of the Monograph:
SEMANTIC BASED PATTERN SEARCH ENGINE

Authors:
Dr P. AJITHA

ISBN: 978-93-86388-47-6
Volume: I
Edition: First

Published by:
Jupiter Publications Consortium
director@jpc.in.net | www.jpc.in.net

Printed by:
Magestic Technology Solutions (P) Ltd, Chennai, India.
info@magesticts.com
www.magesticts.com

Jupiter Publications Consortium
Address: 22/102, Second Street, Venkatesa nagar,
Virugambakkam, Chennai, Tamil Nadu, India.
Website: www.jpc.in.net

ABOUT THE AUTHOR

Dr. P. Ajitha holds a Bachelor of Engineering degree from the Government College of Technology, Coimbatore, and a Master of Engineering degree from AKCE, affiliated with Madurai Kamaraj University. She earned her Ph.D. from Sathyabama Institute of Science and Technology, focusing on Semantic-Based Pattern Search Engines. With over 20 years of teaching experience, Dr. Ajitha has made significant contributions to academia, publishing nearly 50 papers in esteemed national and international journals and conferences. In recognition of her dedication to education, she was honored with the Sarvapalli Dr. Radhakrishnan Best Teacher Award in 2021. Additionally, she has secured grants for several patents and currently serves as the Principal Investigator for the DST-sponsored Agri Innovation Hub, a project aimed at improving the welfare of communities. Her research interests include Data Analytics, Machine Learning, and Text Mining.

QUOTABLE QUOTES

Vannevar Bush: "The human mind... operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts."
This highlights the foundational concept of semantic search, which mimics human associative memory.

Tim Berners-Lee: "The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other related data."
A reminder of the importance of semantics in search engines, where understanding context and relationships between data points enhances search results.

Ray Kurzweil: "Our future is in cognitive computing, and the way forward is through understanding the semantics, the meaning, of the data we interact with."
This emphasizes the role of semantic understanding in the evolution of search technologies.

Peter Norvig: "A search engine should understand the meaning of words, the context they're used in, and the intent of the query."
This directly aligns with the goals of a semantic-based pattern search engine.

Douglas Engelbart: "The better we get at getting better, the faster we will get better."
This reflects the continuous improvement that semantic technologies bring to pattern search and information retrieval.

PREFACE

The rapid advancement of information technology and the ever-growing volume of digital data have necessitated the development of more sophisticated and intelligent search engines. Traditional keyword-based search mechanisms, while effective to a certain extent, often fall short in capturing the nuanced meanings and contextual relationships embedded within the vast expanse of unstructured data. To address these challenges, the concept of a Semantic-Based Pattern Search Engine has emerged, offering a more refined approach to information retrieval that mirrors human cognitive processes.

This monograph delves into the intricate world of semantic search technologies, exploring their theoretical underpinnings, practical implementations, and potential future advancements. It is structured to guide the reader through the various stages of developing a robust semantic search engine, beginning with a comprehensive introduction to the fundamental concepts and culminating in an exploration of advanced applications such as emotion prediction and online opinion mining.

In Chapter 1, we lay the groundwork by introducing the basic principles of semantic search, setting the stage for the more detailed discussions that follow. Chapter 2 defines the objectives and scope of the research, providing a clear framework for the subsequent chapters. The core of this monograph, described in Chapters 3 through 7, focuses on the development of different components of a semantic-based search engine, including concept search, topic search, affective term detection, and emotion prediction engines.

Chapter 8 presents a case study on online opinion mining using a sentiment thesaurus combined with a concept search engine, illustrating the practical applications of the research. Finally, Chapters 9 and 10 offer a reflection on the research findings and propose directions for future enhancements in this evolving field.

This work is intended for researchers, practitioners, and students who are interested in the cutting-edge developments in search technologies. It is my hope that this monograph will contribute to the ongoing discourse in semantic search and inspire further research and innovation in this critical area of study.

Dr. P. Ajitha

ABSTRACT

The evolution of search engines from simple keyword-based systems to more sophisticated semantic-based models marks a significant advancement in the field of information retrieval. This monograph presents a comprehensive study on the development of a Semantic-Based Pattern Search Engine, designed to enhance the accuracy and relevance of search results by understanding the contextual and semantic relationships within data. The research encompasses the creation of various engines, including a concept search engine, topic search engine, affective term detection engine, and emotion prediction engine. Additionally, the application of these technologies in online opinion mining is explored, demonstrating their practical value. The findings of this research contribute to the broader understanding of semantic search and offer valuable insights for future advancements in the domain.

Keywords: Semantic-Based Pattern Search Engine, Concept Search Engine, Topic Search Engine, Affective Term Detection, Emotion Prediction, Online Opinion Mining, Information Retrieval, Semantic Search, Contextual Search

DEDICATION

To my beloved family, whose unwavering support, love, and encouragement have been my greatest source of strength.

To my parents, for their endless sacrifices and for instilling in me the values of perseverance and hard work.

To my spouse, for being my partner in every journey and for your constant belief in my dreams.

To my children, who inspire me every day with their curiosity and joy for life.

To my friends, whose companionship and encouragement have made this journey all the more meaningful.

This work is dedicated to all of you, with heartfelt gratitude.

Dr. P. Ajitha

Contents

CHAPTER 1

INTRODUCTION

History has left huge learning, and technology has captured that into a huge knowledge base. Example search engines like Google searches this huge knowledge base to make our life easier. As technology grows it poses a challenge of mining the knowledge base, which involves lot of scenarios where searching is predominant to make it effective. This calls for search agents to be efficient, scalable and transparent. This research work aims at addressing major challenges faced while categorizing any patterns like text database, image database etc. The challenges in place are the huge dimensionality in categorizing any patterns as well as identifying the relevant dimensions for categorizing the patterns.

Traditionally this has been achieved using numerical vectors, which has left us the above challenges to solve. Given the nature of problem, this research work attempts to address the said challenges by the Semantic based Pattern Classification Mining Model.

CHAPTER 2

OBJECTIVE AND SCOPE OF THE RESEARCH WORK

The main objective of this research work is to provide solution for applications (Emotion, Biometrics), catering Pattern based mining services using pattern search engine model that meets the following criteria.

- Semantic Based Intuitive searching model.
- Narrow search space.
- Efficient, scalable and transparent.
- Find connections between patterns (e.g.: emotions) and affective terms by categorizing the pattern under question in the content under examination.
- Predicts the expected patterns, for example emotions.

CHAPTER 3

DESCRIPTION OF THE RESEARCH WORK

The proposed Semantic based Pattern search Model can be used for various patterns from the likes of Text, Image and other complex patterns like DNA etc… However for the simplicity sake we use the Text as the pattern for our research which in turn is mined using sentence semantics. This model can distinguish between less significant terms and key terms, which hold the concepts intact to represent the sentence meaning.

Pattern Classification mining will extract the most relevant features using Concept Based Model. The major attraction factor for adapting Concept based Pattern classification approach is its inherent transparency, scalability, ease of handling, high precision, accuracy and F-score compared with existing methodologies.

Pattern (Text) classification can be done using Categorization or Clustering. The problem under study is Text categorization and the objective of this research is to overcome the flaws of Text categorization in the conventional methods, by adapting the better

alternative called Semantic Based pattern search engine.

Emotion Prediction and Online Opinion mining are considered here as Use cases for the Text classification and can be achieved using the following building blocks.

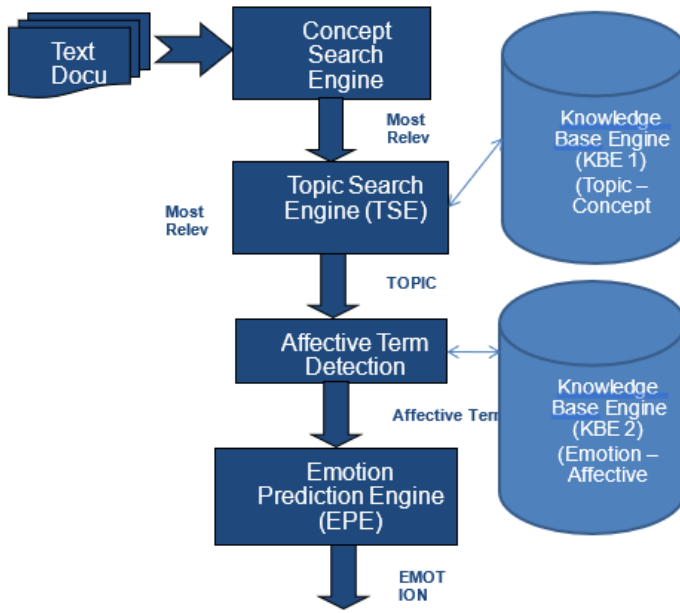

Figure 3.1 Outline of the Proposed Semantic-Based Pattern Search Engine

CHAPTER 4

CONCEPT SEARCH ENGINE

The Concept Search Engine is used to retrieve the most relevant features (concept) from the text document in two phases. The first phase is used to identify the relevant feature from the text document using Concept Based Analysis. The second phase is used to identify the most relevant features (concept) using Lexical Pattern Classification mining.

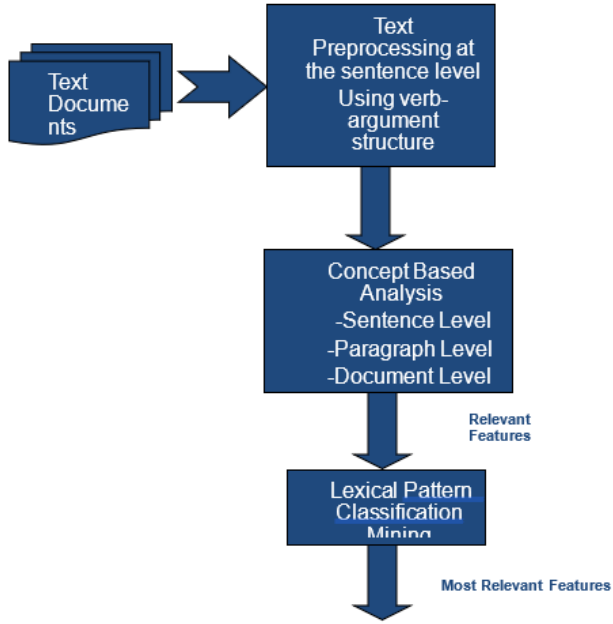

Figure 4.1 Outline of the Proposed Concept Search Engine

The Concept Search Engine is used to analyze the Document at the Sentence level, Document level and

Corpus level to identify the relevant concept that discovers the Topic of the Document.

The Input to the model is the Text document, which consists of collection of sentences. The Semantic Role Labeler identifies the Verb-Noun arguments from the sentences. A single sentence can have many Verb-Noun argument structure based on the availability of information in it. Similarly the same Noun may be used with different Verb which contributes more to the meaning of the sentence. These Verb-Noun arguments can be Terms or Phrases and are called as Concept.

The Concept Term Frequency (CTF) is the number of the occurrence of concept in the Verb-Argument structure of the sentence. The concept which occurs in different Verb-Argument of the same sentence will contribute more to the meaning of the sentence. A concept term (CT) can have different Concept Term Frequency in different sentence of the Document D.

$$\text{CTF} = \frac{\sum_{n=1}^{S_n} ctf_n}{S_n} \quad (1)$$

The overall importance of concept term "CT" weighted to the meaning of its sentences in document "D" is evaluated by calculating the Mean of the Concept