Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 85 (2016) 183 – 189

Procedia Computer Science

International Conference on Computational Modeling and Security (CMS 2016)

# KEYWORD EXTRACTION USING PARTICLE SWARM OPTIMIZATION

D. Sowmya*
PG student
Department of Computer Science and Engineering
Pondicherry Engineering College
Puducherry-605014, India
Sowmyadevaraj93@pec.edu

Dr.J.I.Sheeba
Assistant Professor
Department of Computer Science and Engineering
Pondicherry Engineering College
Puducherry-605014, India
sheeba@pec.edu

**Abstract**

Without formal structure data are those that have no prearranged form or structure and are full of textual data. Typical unstructured systems include emails, reports, telephone or messaging conversations, etc. The main goal of this work is to extract the keywords from a conversation using particle swarm optimization. Keywords are grouped together under their classification and then suggested to the user. In existing work, using diverse keyword extraction, to find topic modelling information, representation of the main topics of transcript and diverse keyword selection. It maximizes the coverage of topics that are automatically recognized in transcript of conversation fragment. Once a set of keywords is extracted, it is clustered according to their user queries and recommended to the user. At the end of result, a single implicit query cannot improve user's satisfaction with the recommended documents. So, swarm intelligence technique is to be applied, it will minimize redundancy in a short list of Keywords and provide accurate query result compared to greedy algorithm.


Peer-review under responsibility of the Organizing Committee of CMS 2016


Peer-review under responsibility of the Organizing Committee of CMS 2016
doi:10.1016/j.procs.2016.05.208

*Keywords:* Keyword Extraction; Recommendation; Clustering; Particle Swarm Optimization.

## 1. Introduction

The aim of keyword extraction from texts is to provide a set of words that are typical to semantic fulfilled of the texts. In the application intended here, keywords are automatically extracted from a conversation and are used to define queries to a just-in-time recommender system. It is thus important that the keyword sets are extracted from the diversity of topics from conversation [1]. Humans are surrounded by an unprecedented wealth of information, available as, databases, documents or multimedia basics even when these are available to the user often do not begin a search , because their uses current activity does not allow them to do so, or because they are not aware that relevant information is available [2].

The main focus is on formulating implicit queries to a just-in-time-retrieval system for use in meeting rooms. In variation to explicit queries that can be made in monetary web search engines, our just-in-time-retrieval system must formulate implicit queries from transcript input, which contains a much larger number of words than a query [3].

For example, in which four people put together a list of items to help them survive in the mountains, a short conversation of 160 seconds contains about 270 words, related to a variety of domains, such as, ‘pistol’, ‘lighter’ or ‘chocolate’. What would be the best accessible 3–6 Wikipedia pages to recommend, and how would a system regulate them?.Once a set of keywords is extracted, it is grouped in order to frame several topically-separated queries, which are run individually, offering better precision than a larger, topically-mixed query [2]. The results are finally grouped into a ranked set before showing them as recommendations to users

Diverse and consistent lists of keywords, which can be recommended to the user of a conversation to accomplish their information needs without, entertain them. These lists bring back regularly by submitting multiple implicit queries derived from the obvious words [4]. Each query is related to one of the topics analyzed in the conversation prior the recommendation, and is acknowledged to a search engine over the Wikipedia. The topic based clustering decreases and the diversity of keywords increases the chances that at least one of the recommended keyword answers a need for information, or can lead to a useful to the user [6].

Pertinence and diversity can be prescribed at three stages: when extracting the keywords into one or several implicit queries or when re-ranking their issues. Then re-ranking results of a single implicit query cannot enhance users’ satisfaction with the recommended keywords. To formulate implicit queries from a text, calculate on word frequency [5] .Others perform keyword extraction by using topic similarity, but do not set a topic diversity constraint. So going to apply swarm intelligence concept, it will improve user satisfaction and easier to cluster the keywords according to their user needed queries [7].

In this framework, it includes identifying the keywords, feature extraction, using particle swarm to reduce