下载的文档为：<https://dl.acm.org/action/doSearch?ConceptID=80&target=ccs-topics>

聚类的时候，就抽取摘要和标题中的关键词即可，如果可以，爬取全部的54554篇文献，然后随机抽取1万个左右聚类；

词库如下：所有的词。

1. Information retrieval:
2. Document representation

Document structure, document topic models, content analysis and feature selection, data encoding and canonicalization, document collection models, ontologies, dictionaries, thesauri.

1. Information retrieval and query processing

Query representation, query intent, query log analysis, query suggestion, query reformulation.

1. Users and interactive retrieval

Personalization, task models, search interfaces, collaborative search

1. Retrieval models and ranking

Rank aggregation, probabilistic retrieval models, language models, similarity measures, learning to rank, combination, fusion and federated search, information retrieval diversity, top-k retrieval in database, novelty in information retrieval.

1. Retrieval tasks and goals

Question answering, document filtering, recommender systems, information extraction, sentiment analysis, expert search, near-duplicate and plagiarism detection, clustering and classification, summarization, business intelligence.

1. Evaluation of retrieval results

Test collections, relevance assessment, retrieval effectiveness, retrieval efficiency, presentation of retrieval results.

1. Search engine architectures and scalability

Search engine indexing, search index compression, distributed retrieval, peer-to-peer retrieval, retrieval on mobile devices, adversarial retrieval, link and co-citation analysis, search with auxiliary database.

1. Specialized information retrieval
2. Structure and multilingual text search

Structured text search, mathematics retrieval, chemical and biochemical retrieval, multilingual and cross-lingual retrieval.

1. Multimedia and multimodal retrieval

Image search, video search, speech/audio search, music retrieval.

1. Environment-specific retrieval

Enterprise search, desktop search, web and social media search