Examples of search algorithms used in IR

Engineering methods 2023/2024

Jaroslav Ertl

Institute of informatics, information systems and software engineering
Faculty of informatics and information technology
Slovak University of Technology in Bratislava

26. november 2023

Table of contents

Introduction

HITS

PageRank

Boolean Search

VSM

Summary and upcoming work

1 HITS

2 PageRank

3 Boolean Search

4 VSM

Introduction

Examples of search algorithms used in IR

Jaroslav Ertl

Introduction

HITS

PageRank

Boolean Search

VSM

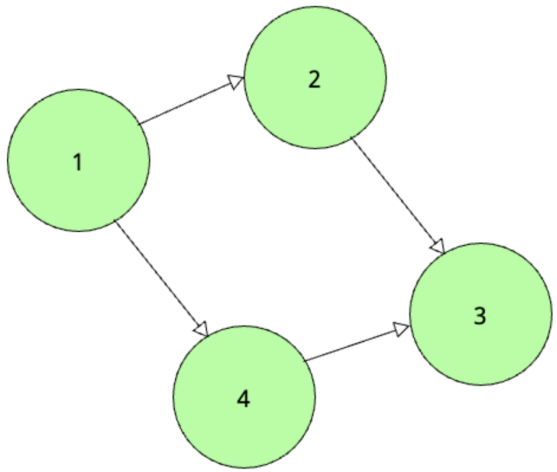
Summary and upcoming work

- What is an algorithm
- Importance of algorithms
- Different types of algorithms

Hyperlink-Induced Topic Search(HITS)

Examples of search algorithms used in IR
Jaroslav Ertl
Introduction
HITS
PageRank
Boolean Search
VSM
Summary and upcoming work

- High-quality pages should link and be linked to by high-quality pages
- Score given to a page
  - Authority
  - Hub
- Algorithm steps
- Problem with bigger networks



PageRank

Examples of search algorithms used in IR

Jaroslav Ertl

Introduction

HITS

PageRank

Boolean Search

VSM

Summary and upcoming work

- Similar to HITS
- PageRank score
- Damping factor
- Algorithm steps

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M_{p_i}} \frac{PR(p_j)}{L(p_j)} \quad 1$$

[1] https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10210566&isnumber=10005208

Boolean Search

Examples of search algorithms used in IR
Jaroslav Ertl
Introduction
HITS
PageRank
Boolean Search
VSM
Summary and upcoming work

- Keyword-based
- Modifying query with operators - NOT, AND, OR
- More operators = less results and more relevant
- Pros
  - Eliminate unwanted page fasts
  - Can be applied to any page
  - Predictable results
- Cons
  - Efficiency relies solely on user

Research

Research with boolean operators done on Google

Query	Number of results
Algorithm	1.37 billion
Search AND algorithm	743 million
Search algorithm AND information	677 million
Search algorithm AND information AND retrieval	133 million
Search algorithm AND information retrieval NOT keyword-based	75.5 million
Search algorithm AND information retrieval NOT keyword-based OR graph-based	28.7 million

Vector Space Model

Examples of search algorithms used in IR
Jaroslav Ertl
Introduction
HITS
PageRank
Boolean Search
VSM
Summary and upcoming work

- Page represented as multidimensional vector
- Terms and weights reflects the importance in space
- Multiple implementations of basic VSM
- TF-IDF(Term Frequency- Inverse Document Frequency)

Summary and upcoming work

Examples of search algorithms used in IR
Jaroslav Ertl
Introduction
HITS
PageRank
Boolean Search
VSM
Summary and upcoming work

- Algorithms play crucial role in IR
- An introduction into the world of algorithms
- Foundation for next generation algorithms
- Finalize the keyword-based algorithms and TF-IDF sections

Examples of search algorithms used in IR
Jaroslav Ertl
Introduction
HITS
PageRank
Boolean Search
VSM
Summary and upcoming work

Thank you for your attention!