

LO1: Search Engine (SE)

Names for different retrieval information (IR) | Document (DK) | Text (TR)

Specific SE

Vertical	One type of data (job search, news search)
Site	Just one site
Custom	narrow search to small set of websites
Enterprise	Corporate intranet w/ collections, metadata, roles, security

Federated search system: Meta search 'glue' system

Federated search (union)	Meta search (filtering hosts)
Search with full-time SE	It itself is not a SE
Agreement obtained	No agreement: many block
Standard query/res. top.	novel and parsing
etc. collaborate	SEs do their own thing

Difficulties of IR

Unstructured text/media	Standard	Semantic	Dynamic/Reactive
-------------------------	----------	----------	------------------

Cloud computing advantages (central IR same)

Virtualization | Many replicas and shards | Scalability for peak

High cost | Inefficient | Inefficient

Timeline of IR: 0 to 3

Web: 1st Gen (1990s: cat, crawler, 1997: hit) | 2nd Gen (1990s: cat, crawler, 1997: hit) | 3rd Gen (1990s: cat, crawler, 1997: hit)

Other things being

search service | cell the software for internet search | Outsource content providers, warehouse by ad (e.g. Yahoo)

Search engine

Do the crawling/indexing/dirty work for clients

Search engine

Page-based algorithm: RES so popup banners

Search engine

Refuse to pay unless unique valuable info

Search engine

by queries, for indexing, for ranking

Search engine

Advanced (query, low) | Address (page content)

Search engine

SE used to need products + new reported, need ad channels

Search engine

SE: SE company help get information, find low rank high

Search engine

Pay-by-impression (low per click = low pay per click)

Search engine

Make web page rank high in algorithm

Search engine

potential legal problems such as lawsuits

Search engine

retrieval needs

Search engine

performance metrics

Search engine

precision: % of relevant in model (noise in second)

Search engine

model: vectors details

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

LO1: Vector space model

doc repr. | query repr. | retrieval function

current class of models

best model | vector space (linear) probabilistic

Common case

$$S_{i,j} = \sum_{k \in X} \frac{1}{2^{d(i,k) + d(j,k) + d(i,j) + 1}}$$

Common case

$$S_{i,j} = \sum_{k \in X} \frac{1}{2^{d(i,k) + d(j,k) + d(i,j) + 1}}$$

Common case (WIS) of (WIS) (WIS)

Recall: size = α (size of set of nodes in path)

size of set of nodes in path = α (size of set of nodes in path)

size of set of nodes in path = α (size of set of nodes in path)

Recall

Page Rank (Google, Hadoop, etc) (WIS) (WIS)

$$D(v_i) = \sum_{j \in P} \frac{u_j}{|P_j(i+1)|}$$

- (1) link structure
- (2) link structure
- (3) link structure
- (4) link structure

$$P(v_i | v_j) = (1-d) + d \sum_{k \in P} \frac{P(v_i | v_k)}{|P_k(i+1)|}$$

Recall: size of set of nodes in path

1. link structure
2. link structure
3. link structure
4. link structure
5. link structure
6. link structure
7. link structure
8. link structure

Recall: size of set of nodes in path

$$P(v_i | v_j) = (1-d) + d \sum_{k \in P} \frac{P(v_i | v_k)}{|P_k(i+1)|}$$

$$P(v_i | v_j) = (1-d) + d \sum_{k \in P} \frac{P(v_i | v_k)}{|P_k(i+1)|}$$

Recall: size of set of nodes in path

Recall: size of set of nodes in path