

4 Associative Search

Dealing with word meanings in information retrieval

■ Example:

Synonyms: Search engine shall find documents with vehicle, auto, ...

Homonyms: The system should recognize that bank has multiple meanings.

■ Dealing with different word meanings

- ◆ Thesaurus: Explicit representation of term relationships
- ◆ Associative Search: Automatic detection of related terms

Associative Search: Taking into account the word „context“

- Two terms are related, if they often occur in the same context
- Context can be
 - ◆ the same document
 - ◆ the same paragraph in a document
 - ◆ in close distance in a document
- The context of a term indicates the meaning of the term
- Example: The term **bank**
 - ◆ in business news, the term **bank** probably occurs together with terms like **dollar**, **assets** or **share**
 - ◆ in literature the term **bank** may occur together with terms like **sit**, **park**, or **tree**



Example: The search engine mindaccess *)

- mindaccess is an associative search engine. For each search the systems provides
 - ◆ a ranked list of documents (like any search engine)
 - ◆ a list of words related to the search term

powered with mindaccess

Verwandte Begriffe

sanwa	fdic	suncook	lloyds	lewiston
assets	dirs	stake	national	bpi

related terms

Themenmix 1..10 von 1500...

result list

Kein Titel vorhanden -
BANK FAILURE OF ... Oklahoma City **bank**, the 4th ... 44th U.S. **bank** failure this ... United Oklahoma **Bank** by United **Bank** of Oklahoma, ... of United **Bank** Shares, Inc., ... oklahoma, fdic, bank, 44th, failure, million, united, city, 13,000, assumption, announces, chartered, newly, accounts, failed
 Kontext
 Schlüsselwörter

Kein Titel vorhanden -
BANK OF CEYLON ... Lanka's largest **bank**, the government-owned **Bank** of Ceylon, ... base, the **bank's** new chairman ... interview. The **bank** aims to ... the Central **Bank** of Ceylon, ... sandaratne, bank, ceylon, rupees, lanka, sri, adopt, export, min, provisions, foreign, nimal, mbd, enlarge, reduce
 Kontext
 Schlüsselwörter

*) mindaccess is a product of Insiders GmbH (<http://www.insiders.de>)



How to find related terms: Principle

	d1	d2	d3
accident	1	0	1
car	1	1	0
cause	0	0	1
crowd	0	0	1
die	1	0	0
drive	0	0	1
four	0	0	1
heavy	1	0	0
injur	0	0	1
more	0	1	0
morning	1	0	0
people	1	0	1
quarter	0	1	0
register	0	1	0
truck	0	0	1
trucker	0	0	1
vehicle	0	1	0
vienna	1	1	1
yesterday	1	0	0

The index matrix can be regarded as two vectors

- ◆ document vectors (columns)
- ◆ **term vectors** (rows)

Information Retrieval:

- ◆ ranking of **document** vectors

Related terms:

- ◆ ranking of **term** vectors

A term t_1 is related to term t_2 , if t_1 and t_2

- ◆ commonly occur in many documents
- ◆ occur solely in few documents

Related terms: Associative Search vs. Thesaurus

- Thesaurus: manually specified term relations (RT = related terms)
 - ◆ Generally valid, based on word meaning
- Associative search: automatically detected term relations
 - ◆ associative Search does **not** consider the **meaning** of terms
 - ◆ Term relations are specific for the considered set of documents. In different document sets the related terms are different
 - ◆ Example:

powered with mindaccess®

Verwandte Begriffe

deutsche	680,5	ce telem	kooknin	berliner
übernahme	wgz	xia	bankgesellschaft	hana

powered with mindaccess®

Verwandte Begriffe

sanwa	fdic	suncook	lloyds	lewiston
assets	dlrs	stake	national	bpi

How to use term relations: Query refinement

- Clicking on related terms extends the query (and the system re-computes the related terms)
- Useful for *homonyms* to restrict answer set to correct context

The screenshot shows the MindAccess search interface with the query 'bank assets'. Below the search bar, there is a table of related terms: 'savings', 'dlrs', 'lewisston', 'Federal', 'fdic', 'lhrift', 'fhbb', 'home', 'mln', 'banks'. Below this, there are two search results. The first result is for 'BANK FAILURE OF ... Oklahoma City bank, the 44th ... 44th U.S. bank failure this ... United Oklahoma Bank by United Bank of Oklahoma, ... of United Bank Shares, Inc., ... Oklahoma, fdic, bank, 44th, failure, million, united, city, 13,000, assumption, announces, chartered, newly, accounts, failed'. The second result is for 'BANK OF CEYLON ... Lanka's largest bank, the government-owned Bank of Ceylon, ... base, the bank's new chairman ... interview. The bank aims to ... the Central Bank of Ceylon, ... sandaratne, bank, ceylon, rupees, lanka, sri, adopt, export, mln, provisions, foreign, nimal, mbsl, enlarge, reduce'. Both results show 'Kein Titel vorhanden' and 'Schlüsselwörter'.

How to use term relations: Query Adaptation

- Often words occur in text together with *synonyms* or words which have nearly the same meaning
- If relevant documents are not found in the first N documents, exchange the query term(s) with related terms
 - ◆ other documents with the related terms may be found

The screenshot shows the MindAccess search interface with the query 'car'. Below the search bar, there is a table of related terms: 'production', 'seats', 'fabriquiry', 'sales', 'output', 'pet', 'chrysler', 'hyundai', 'motor'. Below this, there are three search results. The first result is for 'CAR OUTPUT DECLINE ... amc, car, 2,978, BRASEL, units, declined'. The second result is for 'CHRYSLER ... PUNCT TEXT Chrysler Canada Ltd, ... owned by Chrysler Corp. Chrysler for sale ... 1987, and Chrysler Canada plans ... model, colt, rmc, door, subcompact, colts, import, subcompact, hatchback, units, sedan, sale, 9,000'. The third result is for 'Chrysler Corp said ... North ... truck, pet, north, year, output, fah, fd, Schlüsselwörter'. Callouts indicate: 'Query with car finds 280 documents', 'Refinement with related term "chrysler" reduces to 46 documents', and 'Using chrysler alone finds 128 document, i.e. new documents are found'.

Associative Search: User Feedback (2)

powered with mindaccess

Verwandte Begriffe

erm	park	veloam	zirkus	tage
caracas	carstica	romastron	schische	erom

neue Themen

1-10 von 150...

"Schumi" voted athlete of the year by French newspaper =
Tiger Woods in ...
 champion, newspaper, schumacher, voted, schumacher, year, french, triumphed, golfer, schumi, Pequeño, athlete, footballer, tennis, skiing

Kontext Schlüsselwörter

NEWS FEATURE: Masters more than usual "race for second" behind Woods =
Tiger Woods: A ... defending champion **Tiger** Woods: But ... faller, "He (**Tiger**) is the ..."
 augusta, woods, masters, midweek, dural, feel, course, year, putt, putting, confidence, finishes, winning, victory, usual

Kontext Schlüsselwörter

ROUNDUP: Weir shoots bogey on first playoff hole to win Masters: Eds: adds detail ...
Tiger Woods: failed ...
 weir, mullace, hole, putt, masters, playoff, win, bogey, par, canadian, championship, green, shooting, day, title

Kontext Schlüsselwörter

Woods, Ronaldo and Schumacher shortlisted for Laureus awards =
Tiger Woods aims ...
 award, laureus, ronaldo, olympic, woods, champion, shortlisted, world, schumacher, golfer, winner, tennis, kostelc, star, champions

Kontext Schlüsselwörter

NEWS FEATURE: Woods unstoppable in march to Masters history =
Tiger Woods gave ... I think **Tiger** will be ... playing with **Tiger**. Even playing ... an average **Tiger** performance. World ... to catch **Tiger**, herats we ...
 woods, goosen, holes, nicklaus, masters, bogeys, shot, 1-over, 1-uder, proved, round, championship, major, third, three

Kontext Schlüsselwörter

Kanadier Weir gewann US-Masters - Woods 15. - Langer ausgeschieden =
Tiger Woods: der ...
 weir, schlägen, augusta, mullace, masters, woods, steehen, kanadier, par, traditionturners, eingelocht, langer, 75er, rang, putt

Kontext Schlüsselwörter

Applying User Feedback:

- terms in the relevant documents are preferred
- documents with preferred terms have a higher ranking

Example:

- The first documents only contain documents with „tiger“ in the context of „woods“

Associative Search: User Feedback (3)

Willkommen: K.Hinkelmann

Erweiterte Suche

Globale Suchbereiche

- Suchbereiche
- Sprache
- NewsSparten
- Jahrgang

Eigene Suchbereiche

- neue Themen
- tiger woods sport
- Meine Ablage
- Webseitenkanäle

Dieses Feedback als eigenes Thema speichern

Dieses Feedback als eigenes Thema speichern:

- "Schumi" voted athlete of the year by French newspaper =
- NEWS FEATURE: Masters more than usual "race for second" behind Woods =
- Drei Sibirische Tiger in französischem Zirkus geboren =
- Satz Magier Roy kämpft nach Tiger-Angriff auf der Bühne um sein Leben Von Christiane Oelrich, dpa =
- FEATURE: Tiger hunter turns conservationist in Cambodia By Bronwyn Sloane, dpa Eds: epa photos available =
- Ein Jahr nach dem Tigerangriff - Roy Horn feiert 60. Geburtstag mit den Bildern FKM12/13 vom 28.9.) Von Barbara Munkler, dpa =

Themenname: tiger woods sport

Speichern Abbrechen

weir, mullace, hole, putt, masters, p, day, title

Woods, Ronaldo and Schuma-
Tiger Woods aims ...
 award, laureus, ronaldo, olympic, tennis, kostelc, star, champions

NEWS FEATURE: Woods unst-
Tiger Woods gave ... I think **Tig**
Tiger performance. World ... to c
 woods, goosen, holes, nicklaus, ma
 championship, major, third, three

Kanadier Weir gewann US-M-
Tiger Woods: der ...
 weir, schlägen, augusta, mullace, r
 eingelocht, langer, 75er, rang, putt

Storing Feedback:

- Feedback can be stored as a topic
- Topics can be applied as filter in later search
- Topics may be used to sort/rank search results thematically