Jul – May 2022
IT458

Capturing User Relevance

Query Expansion Techniques

Relevance Feedback v/s Query Expansion

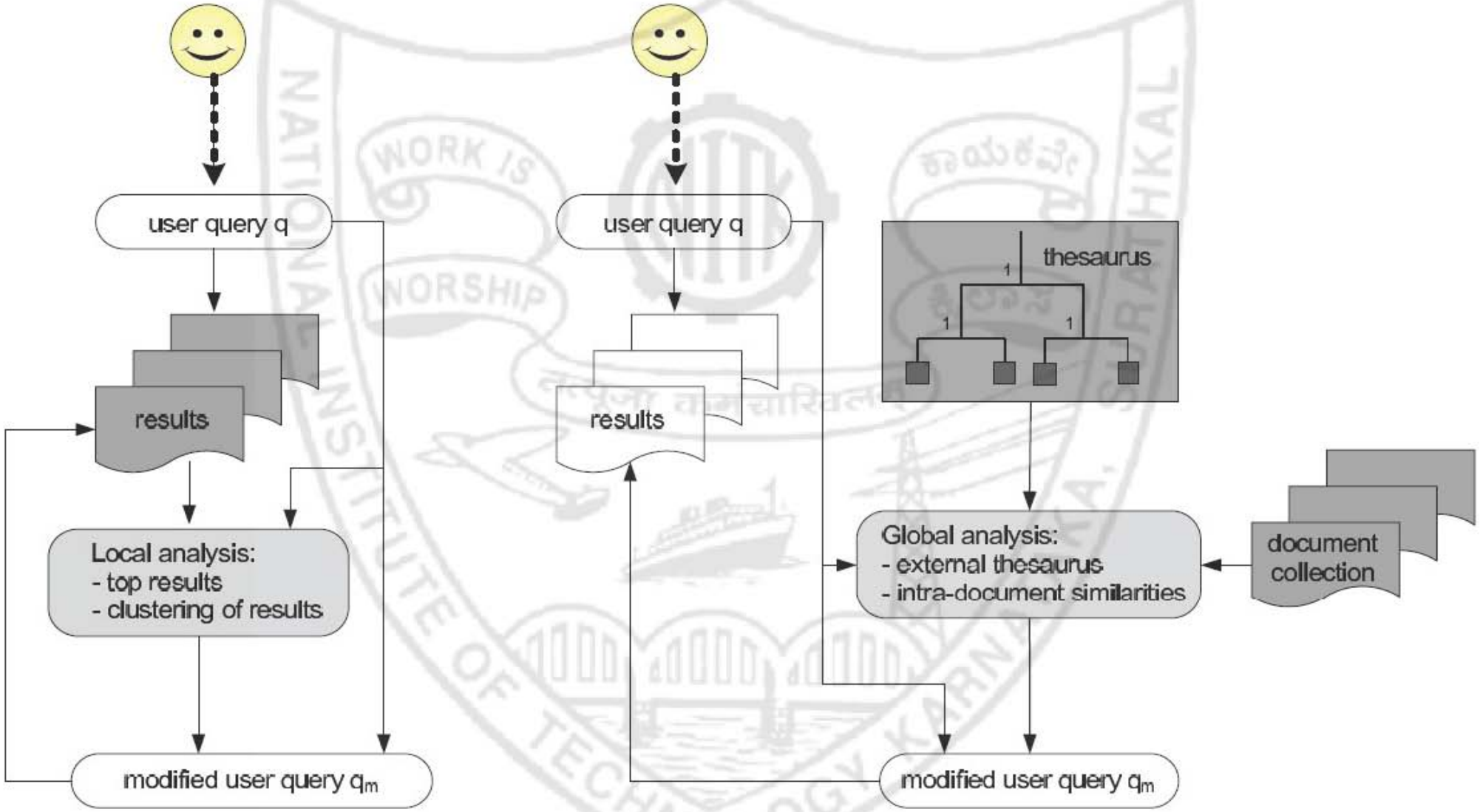
- In relevance feedback, users give *(or system generates)* additional input (relevant/non-relevant) on documents
  - used to **reweight terms** in the documents

- In query expansion, users give *(or system generates)* additional input (good/bad search term) on words or phrases.
  - used to **expand** the query.

Query expansion

- Two basic approaches for compiling implicit feedback information:
  - **local analysis**
    - Derive feedback info from top ranked documents in the generated result set.
  - **global analysis**
    - Derive feedback info from external sources also.
      - Example: a thesaurus.

Query expansion



Query Expansion

- **Local Analysis:(dynamic)**
  - Analysis of documents in result set
- **Global Analysis (static; of all documents in collection)**
  - Controlled vocabulary
  - Manual thesaurus generation
  - Automatically derived thesaurus
  - Refinements based on query log mining

Local Analysis based Query Expansion

Automatic Local Analysis

- At query time, dynamically determine similar terms based on analysis of top-ranked retrieved documents.
  - analysis on only the "local" set of retrieved documents for a specific query.

- Avoids ambiguity by determining similar (correlated) terms only within relevant documents.
  - "Apple computer" → "Apple computer Powerbook laptop"

Automatic Local Analysis

- Only expand query with terms that are similar to all terms in the query.

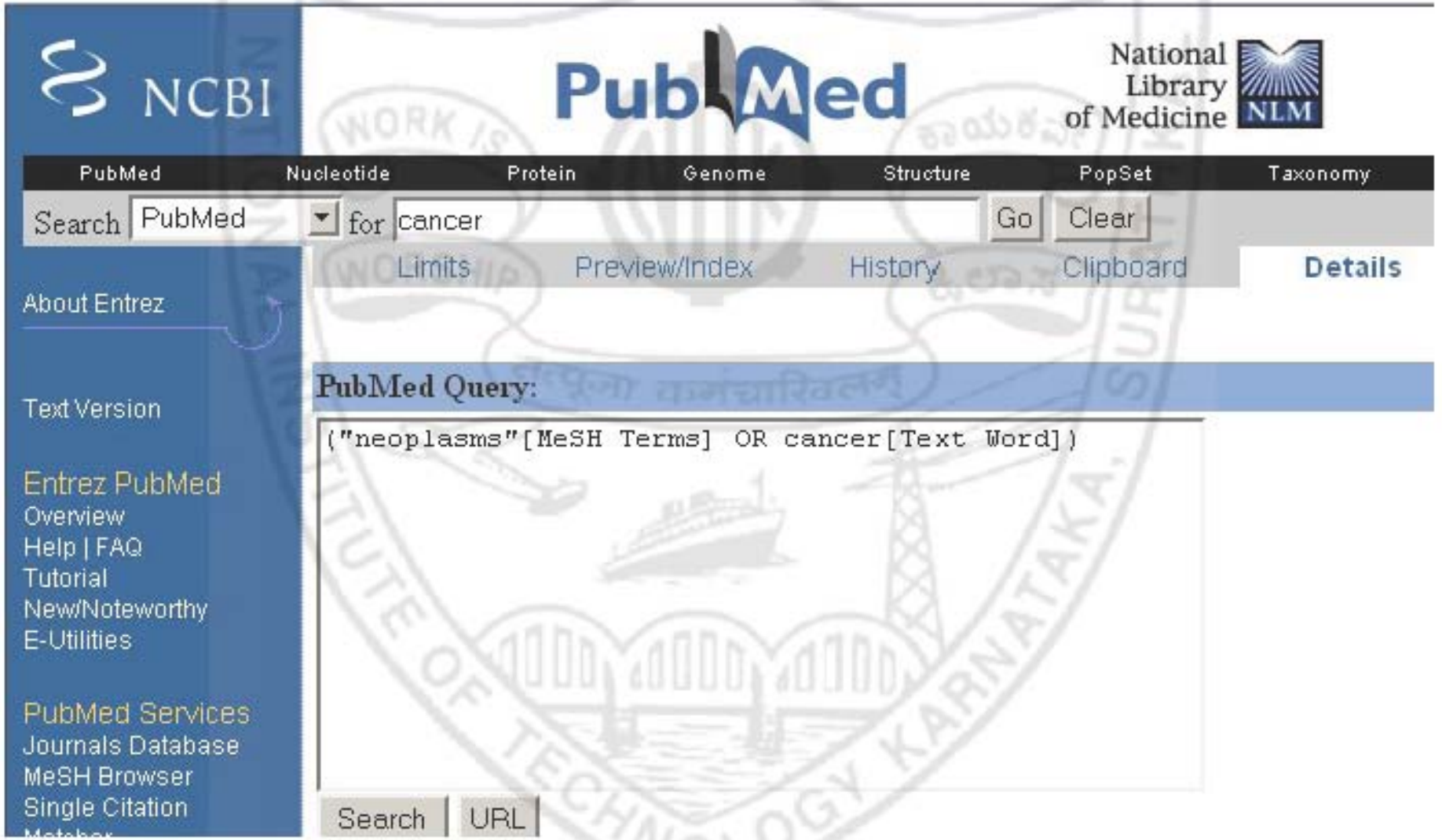
$$sim(k_i, Q) = \sum_{k_j \in Q} c_{ij}$$

  - "*fruit*" not added to "*Apple computer*" since it is far from "*computer*."
  - "*fruit*" added to "*Apple pie*" since "*fruit*" close to both "*apple*" and "*pie*."

- Uses more sophisticated term weights (instead of just frequency) when computing term correlations.

Global Analysis based Query Expansion Approaches

Controlled Vocabulary



Thesaurus-based Query Expansion

- For each term, t, in a query, expand the query with synonyms and related words of t from the thesaurus
  - E.g. feline → feline cat
- Features:
  - Does not require user input.
  - May weight added terms less than original query terms.
- Implementation Example : WordNets

Princeton WordNet

- A large lexical database, or "electronic dictionary" of semantic relationships between English words (About 144,000 words)
- Nouns, adjectives, verbs, and adverbs grouped into synonym sets called **synsets.**
- Synsets are interconnected
  - Bi-directional arcs express semantic relations
  - Result: large semantic network (graph)

WordNet Relationships

- Captures two fundamental, universal properties of human language:
  - **Polysemy**
    - The coexistence of many possible meanings for a word or phrase of a language.
  - **Synonymy**
    - Word/phrase that means exactly or nearly the same as another word/phrase of a language.

WordNet Relationships - Polysemy

- One word form expresses multiple meanings
  - *E.g. table* (as a noun or verb) is eightfold polysemous, i.e., has eight senses

WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: table Search WordNet

Display Options: (Select option to change) Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) **table**, tabular array (a set of data arranged in rows and columns) *"see table 1"*
- S: (n) **table** (a piece of furniture having a smooth flat top that is usually supported by one or more vertical legs) *"it was a sturdy table"*
- S: (n) **table** (a piece of furniture with tableware for a meal laid out on it) *"I reserved a table at my favorite restaurant"*
- S: (n) mesa, **table** (flat tableland with steep edges) *"the tribe was relatively safe on the mesa but they had to descend into the valley for water"*
- S: (n) **table** (a company of people assembled at a table for a meal or game) *"he entertained the whole table with his witty remarks"*
- S: (n) board, **table** (food or meals in general) *"she sets a fine table"; "room and board"*

Verb

- S: (v) postpone, prorogue, hold over, put over, **table**, shelve, set back, defer, remit, put off (hold back to a later time) *"let's postpone the exam"*
- S: (v) **table**, tabularize, tabularise, tabulate (arrange or enter in tabular form)

WordNet Relationships - Synonymy

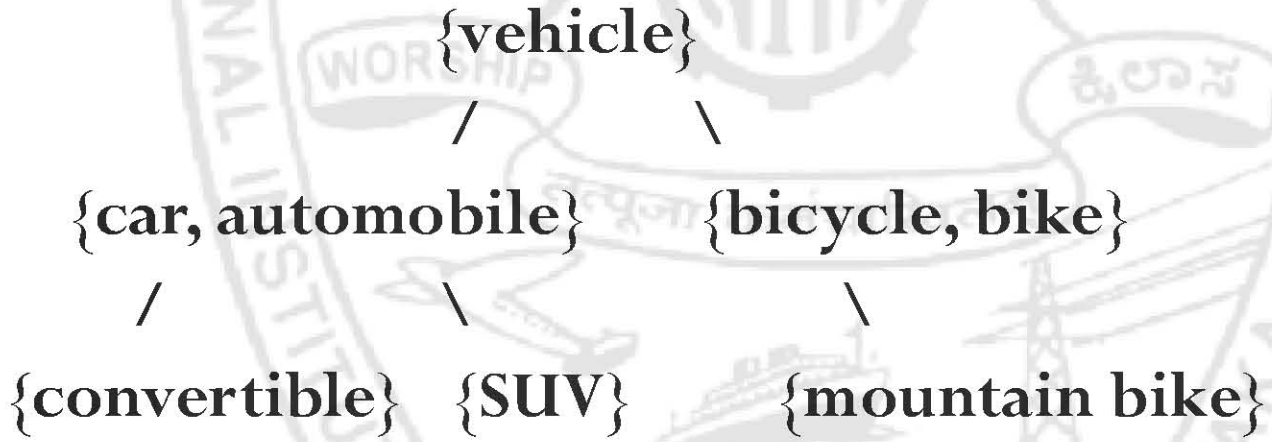
- Synonymous (denotationally equivalent) words are grouped into unordered sets of synonyms ("synsets")
  - *{hit, beat, strike}*
  - *{big, large}*
  - *{queue, line}*
- By definition, each synset expresses a distinct meaning/concept .
  - Each word form-meaning pair is unique.

Some WordNet stats

Part of speech	Word forms	Synsets
noun	117,798	82,115
verb	11,529	13,767
adjective	21,479	18,156
adverb	4,481	3,621
total	155,287	117,659

WordNet Relationships: Hypo/Hyper-nymy

- Relates more/less general concepts
- Creates hierarchies, or "trees", can have up to 16 levels



- **"A car is a kind of vehicle" <=> "The class of vehicles includes cars, bikes"**

WordNet Relationships - Hypo/Hyper-nymy

- Transitivity:

A car is a kind of vehicle

A SUV is a kind of car

=> A SUV is a kind of vehicle

WordNet Relationships - Meronymy/Holonymy

- **part-whole relation**

{car, automobile}
|
{engine}
/ \
{spark plug} {cylinder}

=> "An engine has spark plugs"
=> "Spark plugs and cylinders are parts of an engine"

- **Inheritance:**

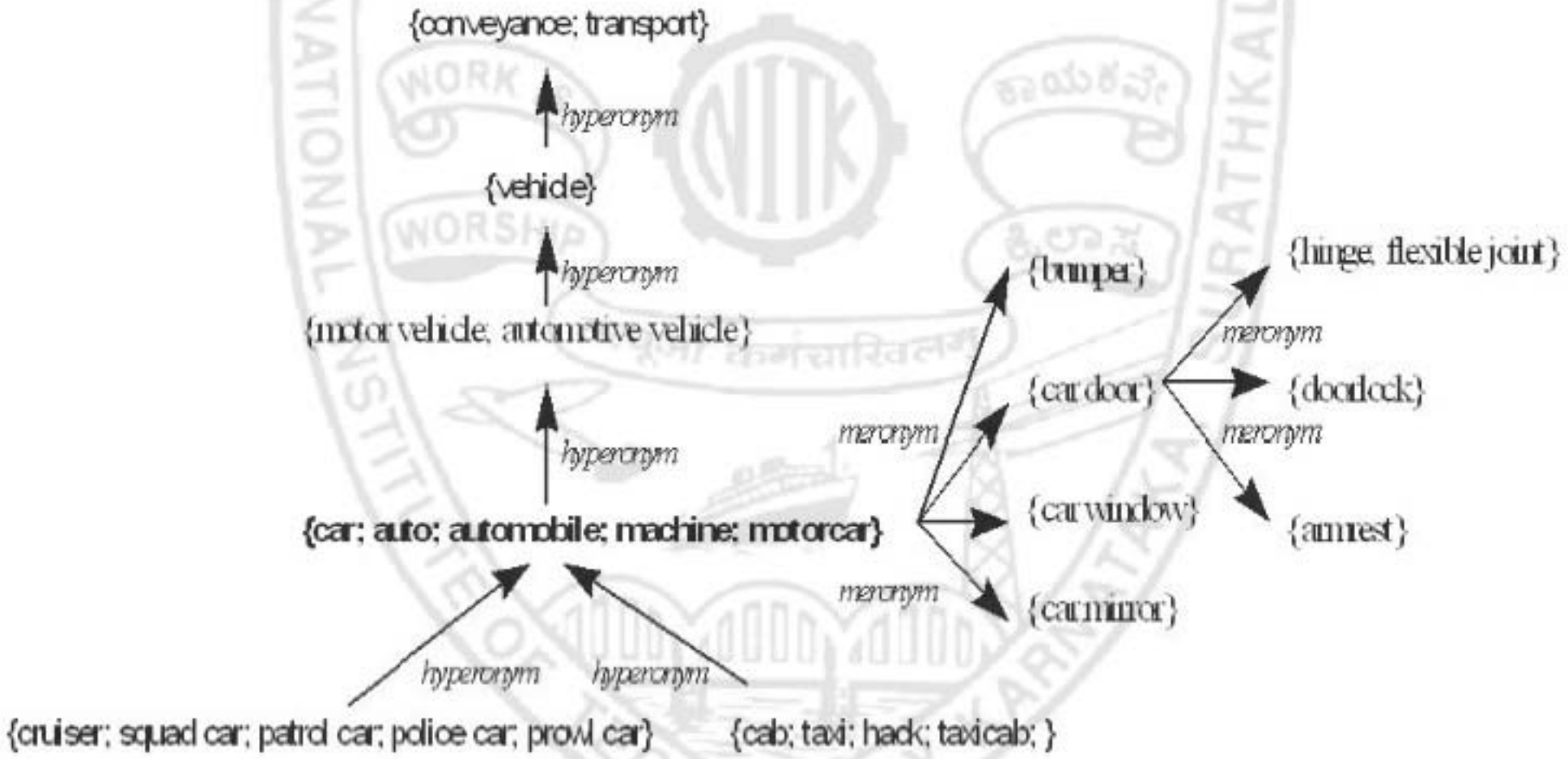
A finger is part of a hand
A hand is part of an arm
An arm is part of a body

=>a finger is part of a body

WordNet Synset Relationships

- Antonym: front → back
- Attribute: benevolence → good (*noun* to *adjective*)
- Similar: unquestioning → absolute
- Cause-effect: kill → die
- Entailment: breathe → inhale
- Holonym: chapter → text (*part-of*)
- Meronym: computer → cpu (*whole-of*)
- Hyponym: tree → plant (*specialization*)
- Hypernym: apple→ fruit (*generalization*)

Structure of WordNet (Nouns)



WordNet based Query Expansion

- WordNet Query Expansion Process -
  - Add synonyms in the same synset.
  - Add hyponyms to add specialized terms.
  - Add hypernyms to generalize a query.
  - Add other related terms to expand query.

Automatic Thesaurus Generation

- Attempt to generate a thesaurus automatically by analyzing the collection of documents **(Domain Modeling)**

- Two main approaches
  - Co-occurrence based
    - co-occurring words are more likely to be similar
  - Shallow analysis of grammatical relations
    - E.g. entities that are *grown*, *cooked*, *eaten*, and *digested* are more likely to be **food** items.

Automatic Thesaurus Generation

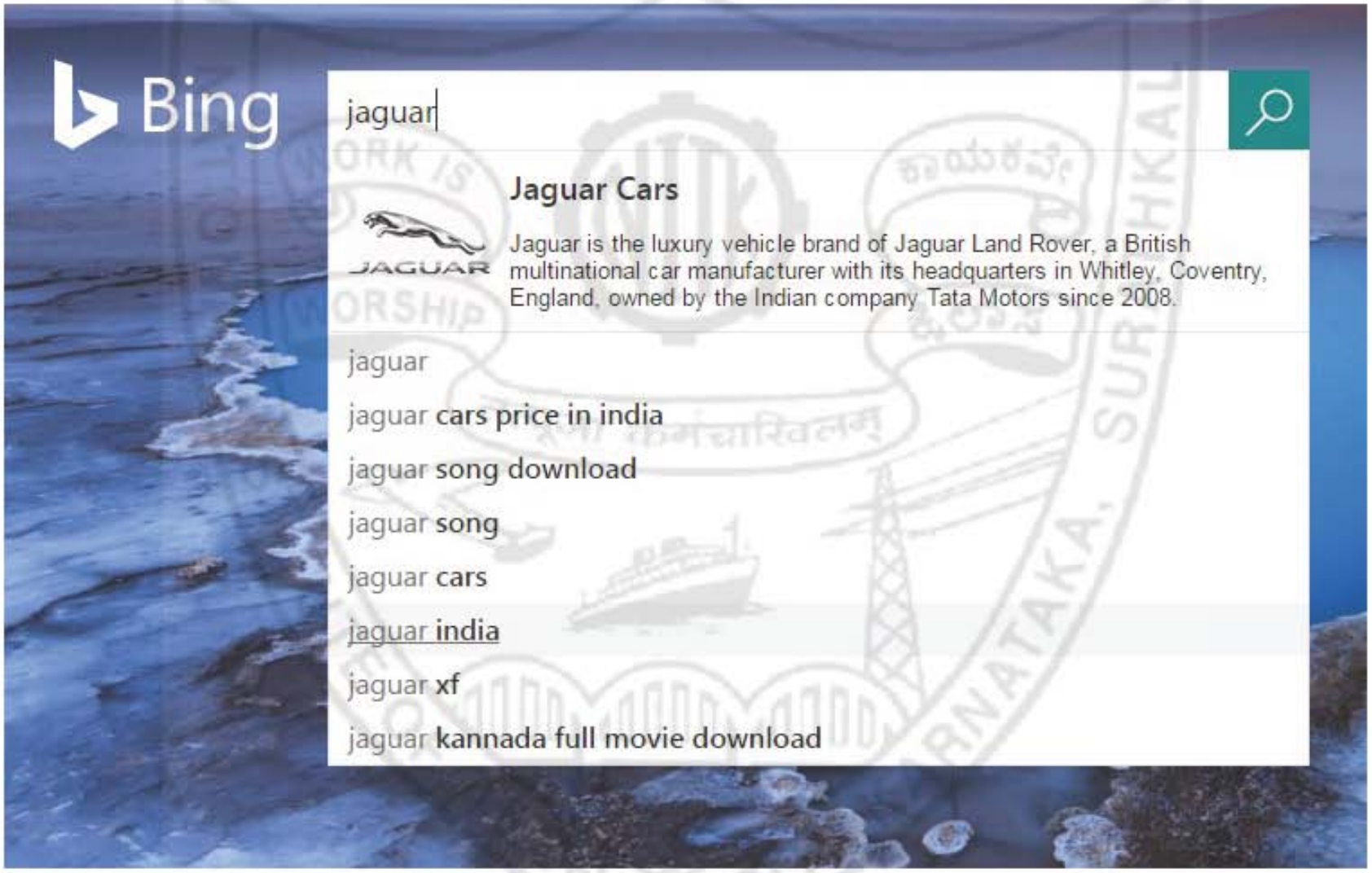
- Attempt to generate a thesaurus automatically by analyzing the collection of documents **(Domain Modeling)**

- Two main approaches
  - Co-occurrence based
    - co-occurring words are more likely to be similar
  - Shallow analysis of grammatical relations
    - E.g. entities that are *grown*, *cooked*, *eaten*, and *digested* are more likely to be **food** items.

* Co-occurrence based is more robust, grammatical relations are more accurate.

Refinements based on Query Log mining

Refinements based on Query Log mining



Global vs. Local Analysis

- Global analysis
  - requires intensive term correlation computation only once at system development time.
- Local analysis
  - requires intensive term correlation computation for every query at run time
  - number of terms and documents is less than in global analysis.

* Local analysis gives better results.

Query Expansion: Summary

- Query expansion is often effective in increasing recall.
  - Not always with general thesauri
  - Fairly successful for subject-specific collections
- In most cases, precision is decreased, often significantly.
- Overall, not as useful as relevance feedback; may be as good as pseudo-relevance feedback

More reading

- Carpineto, C., & Romano, G. (2012). A survey of automatic query expansion in information retrieval. *Acm Computing Surveys (CSUR)*, *44*(1), 1-50.

- Miller, G., "WordNet: a lexical database for English." *Communications of the ACM* 38, no. 11 (1995): 39-41.

- Xu, J., Croft, W.B. (1996): Query Expansion Using Local and Global Document Analysis, in SIGIR 19: 4-11.