



IR System Design

Fundamental Concepts

IR Systems

FUNDAMENTAL NOTIONS...

A “document”...

- ▶ A document is a **coherent** passage of **free text** –
 - ▶ “Coherent” → **is about related topics**
 - ▶ “Free” → **natural, written language**

- ▶ Examples:
 - ▶ Newspaper article
 - ▶ Scientific article
 - ▶ Dictionary entry
 - ▶ Web page
 - ▶ Email message...



A “Document Collection”...

- ▶ A set of documents –
 - ▶ Also known as **corpus/corpora**
 - ▶ Usually, all documents within a collection are added w.r.t some criterion.

- ▶ Examples:
 - ▶ Wikipedia
 - ▶ The articles covered by Google News
 - ▶ The Web ..



An “Information Need”...

- ▶ the topic about which the user desires to know more about..
 - ▶ “Refers to an individual's hidden cognitive state” → Ref: R. Taylor, ‘The process of asking questions’, 1962
 - ▶ Often ill-defined and vague.
 - ▶ Depends on what the user knows and doesn't know!



An “Information Need”...

- ▶ Examples:
 - ▶ *“I would like to know about how Google uses semantics and machine learning to improve search”*
- ▶ Translate *need* into a *query*
 - ▶ Can be person/time/context-specific...
 - ▶ Specific keywords used can affect search results..



A “Query”...

- ▶ What the user submits to the system in an attempt to convey their information need.

- ▶ Stated using a formal query language
- ▶ Usually a sentence/phrase/set of keywordslist that consist of search terms.

A “Query”...

- ▶ Example: “I would like to know about how Google uses semantics and machine learning to improve search”
 - ▶ User 1: google search working
 - ▶ User 2: “google semantic search”
 - ▶ User 3: “machine learning at google”
 - ▶ ... “google webpage ranking algorithm”

A “Query”...

- ▶ More formal/structured representations can also be used.
 - ▶ Typically adopted by experienced/expert users

The image displays two side-by-side Google search results pages.

Top Screenshot: The search query is "Hummingbird IN Search BUT NOT bird". The results include:

- Google Hummingbird - Wikipedia**
en.wikipedia.org › wiki › Google_Hummingbird
Hummingbird is the codename given to a significant algorithm change in Google Search in 2013. Its name was derived from the speed and accuracy of the ...
- Google's Hummingbird Update: How It Changed Search**
www.searchenginejournal.com › Google Algorithm History
12-Apr-2022 · Learn what Google Hummingbird really was and see how Googlers and SEO experts explain the impact it eventually had on search.

Bottom Screenshot: The search query is "nitk.ac.in:IT458". The results include:

- IT458 | Department of Information Technology - NITK Surathkal**
infotech.nitk.ac.in › course › it458
IT458. Course Name: Information Retrieval (IT458). Programme: B.Tech (IT). Category: Programme Specific Electives (PSE). Credits (L-T-P): (3-0-2) 4 ...
- Summary of IT458 - Information Retrieval - Courses@IRIS**
courses.iris.nitk.ac.in › course › info
Home · Courses · Information Technology · IT458 - 28325; Summary. IT458 - Information Retrieval. Teacher: Sowmya Kamath - IT458 - 28325.

“Information Need” to “Query”

Taylor's four levels of question formation

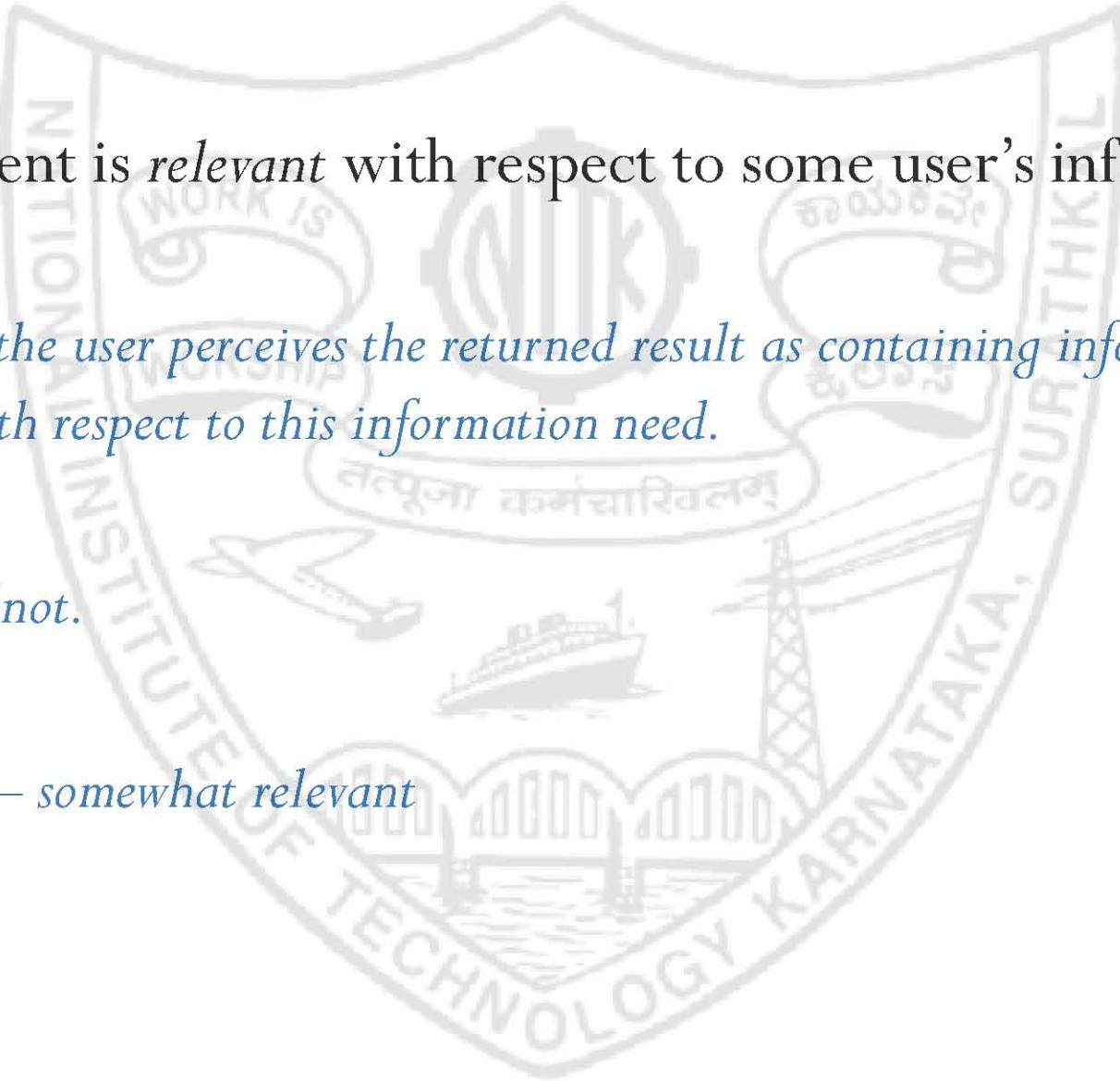
Q1 The actual but unexpressed need for information
(the visceral need)

Q2 The conscious, within-brain description of the need
(the conscious need)

Q3 The formal statement of the need
(the formalized need)

Q4 The question as presented to the information system
(the compromised need)

“Relevance”

- 
- ▶ A document is *relevant* with respect to some user's information need.
 - ▶ Yes - *if the user perceives the returned result as containing information of value with respect to this information need.*
 - ▶ No – *if not.*
 - ▶ Maybe – *somewhat relevant*

“Relevance”

- ▶ A document is *relevant* with respect to some user’s information need.
 - ▶ Yes - *if the user perceives the returned result as containing information of value with respect to this information need.*
 - ▶ No – *if not.*
 - ▶ Maybe – *somewhat relevant*
- * Usually assumed to be a binary concept, but could also be graded.

"Relevance"

"Relevance appears to be a subjective quality, unique between the individual and a given document, supporting the assumption that relevance can only be judged by the information user."

- *Miranda Pao, Professor of Information and Library Studies,
Univ. of Michigan.*

Defining relevance...

- ▶ Two faces of relevance
-

**System-defined
relevance**

**User-defined
relevance**

Defining relevance...

- ▶ Two faces of relevance

System-defined
relevance

vs.

User-defined
relevance

✓ *Objective*

✓ *Subjective*

Defining relevance...

- ▶ Two faces of relevance

System-defined
relevance

vs.

User-defined
relevance

- ✓ *Objective*
- ✓ *Often topical.*

- ✓ *Subjective.*
- ✓ *Situational.*

Defining relevance...

- ▶ Two faces of relevance

System-defined
relevance

vs.

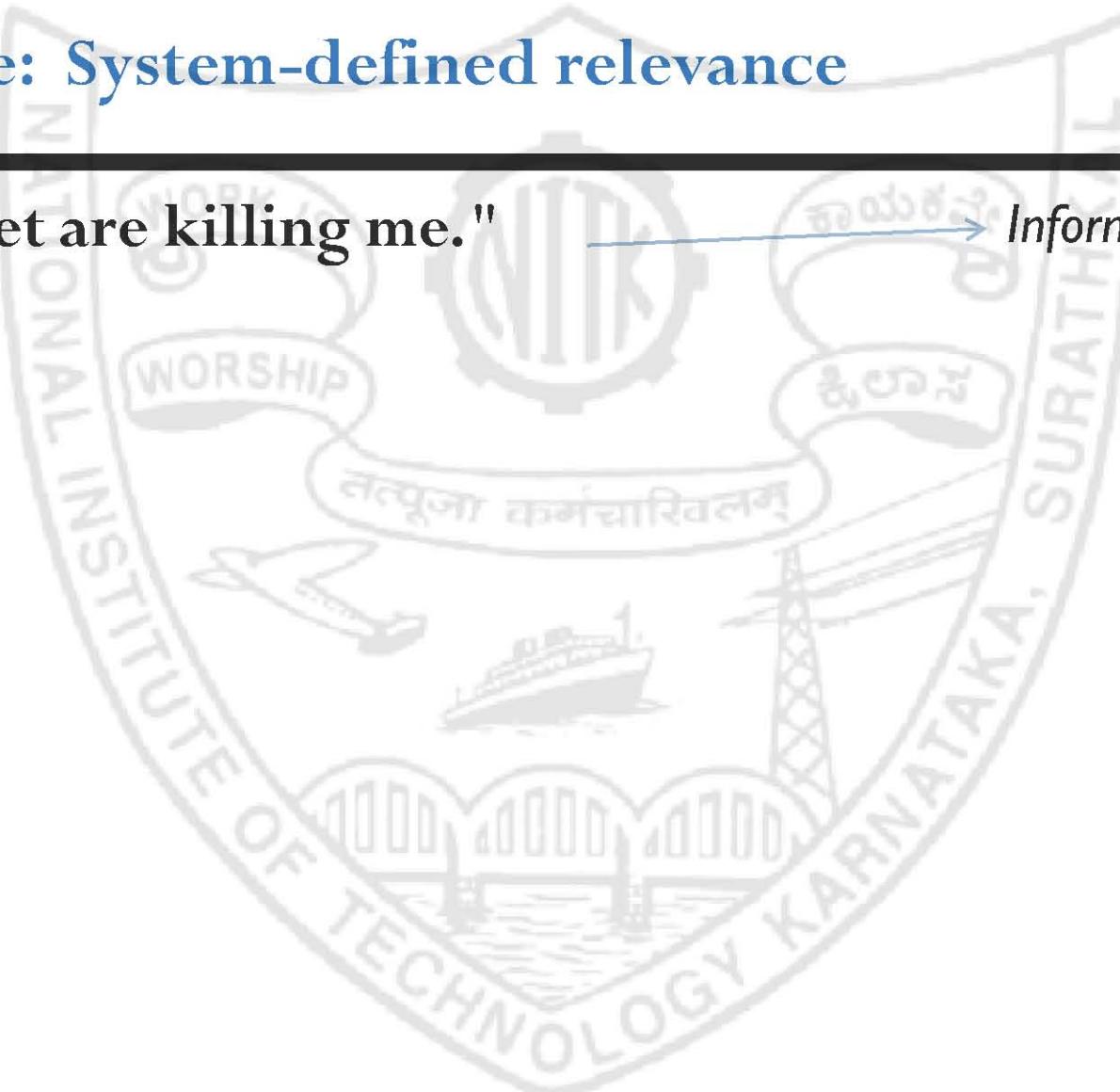
User-defined
relevance

- ✓ *Objective*
- ✓ *Often topical.*
- ✓ *Does it match the query?*
- ✓ *Subjective.*
- ✓ *Situational.*
- ✓ *Is it useful to me?*

Defining relevance...

► Example: System-defined relevance

"My feet are killing me." → *Information need*



Defining relevance...

► Example: System-defined relevance

"My feet are killing me."

Information need

Healthy feet

Query

Defining relevance...

► Example: System-defined relevance

"My feet are killing me."

Information need

Healthy feet

Query

The healthy look of lumber .. 90%
industry measures in terms of
cubic feet of lumber produced ..

*The document deemed
most relevant by a simple
word-matching system.*

Defining relevance...

► Example: User-defined relevance

"My feet are killing me." → *healthy feet*

The healthy look of lumber .. 90% industry measures in terms of cubic feet of lumber ...

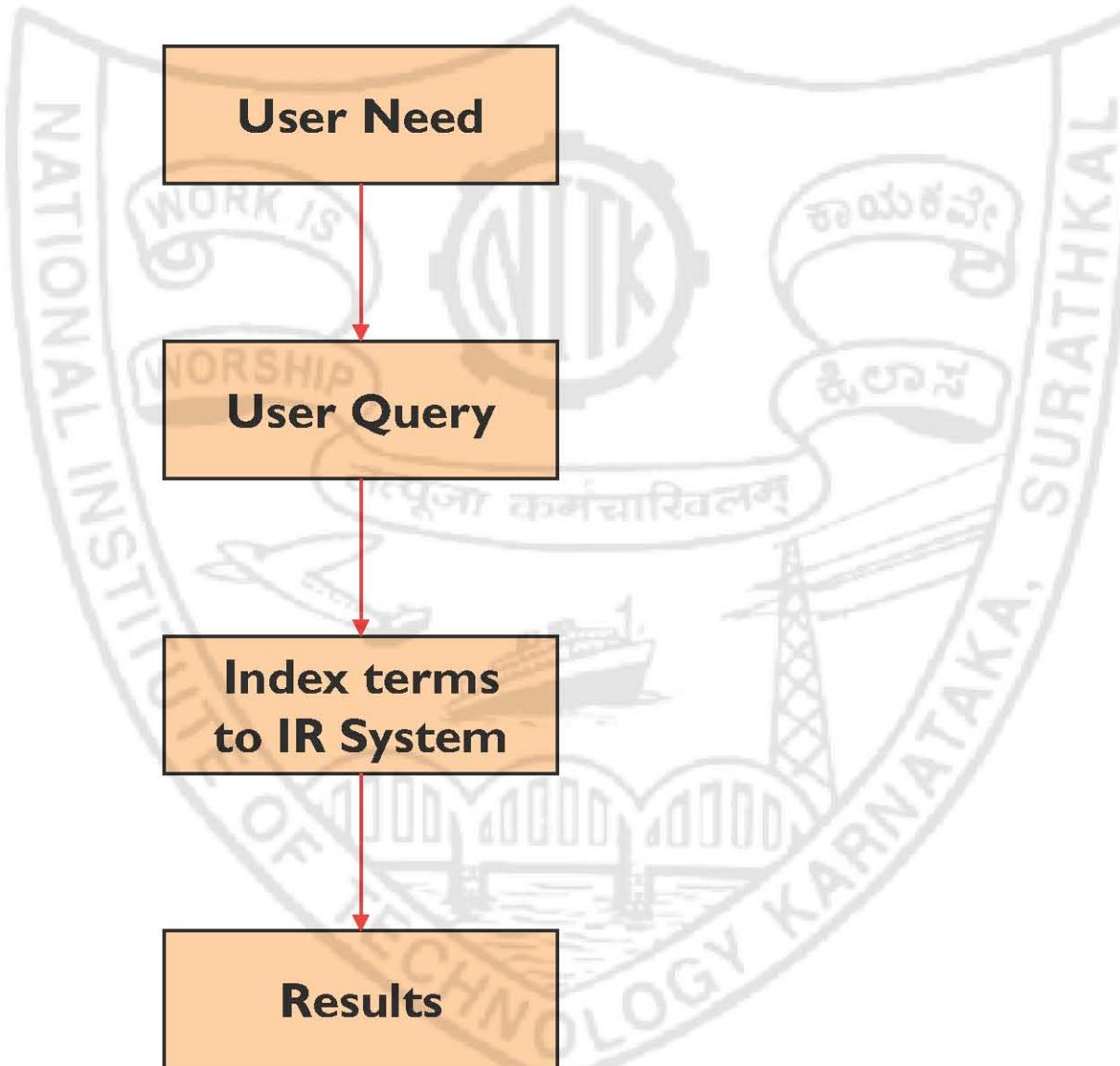
Word-matching system

Soothing remedies for aching feet

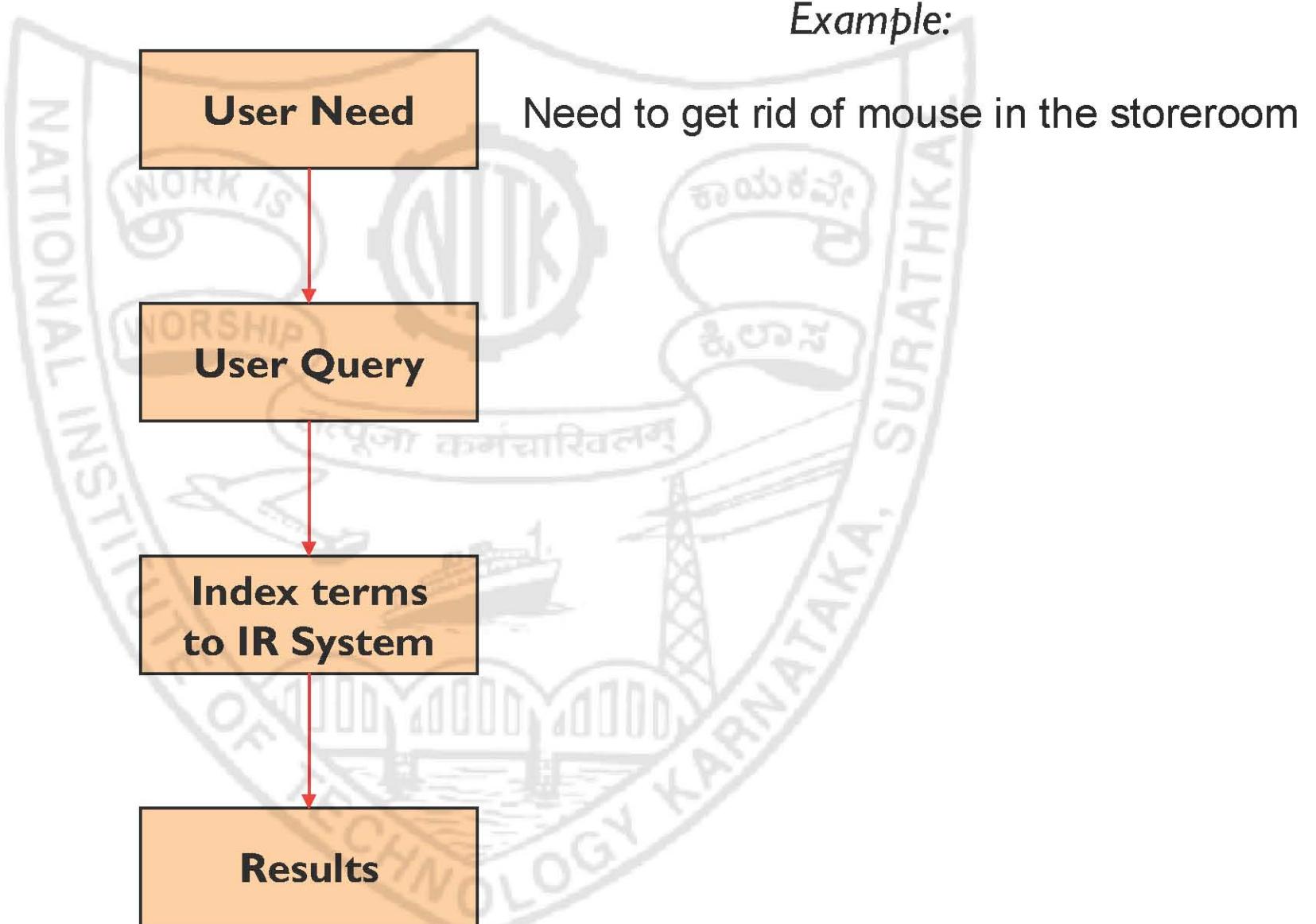
A more intuitive,
context-aware
system

Controlling the body by controlling the mind--
meditative techniques for dealing with pain

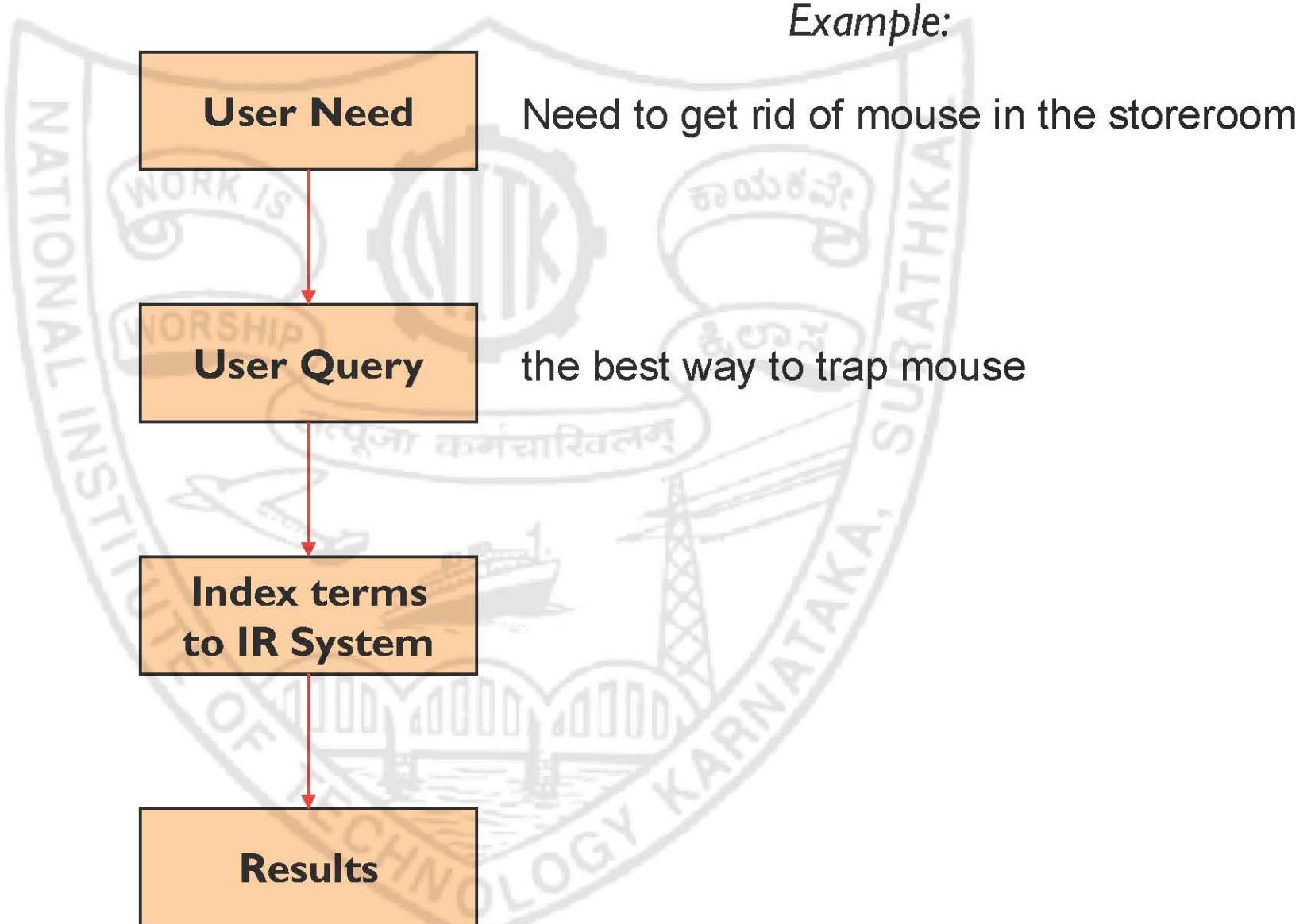
Why Don't Users Get What They Want?



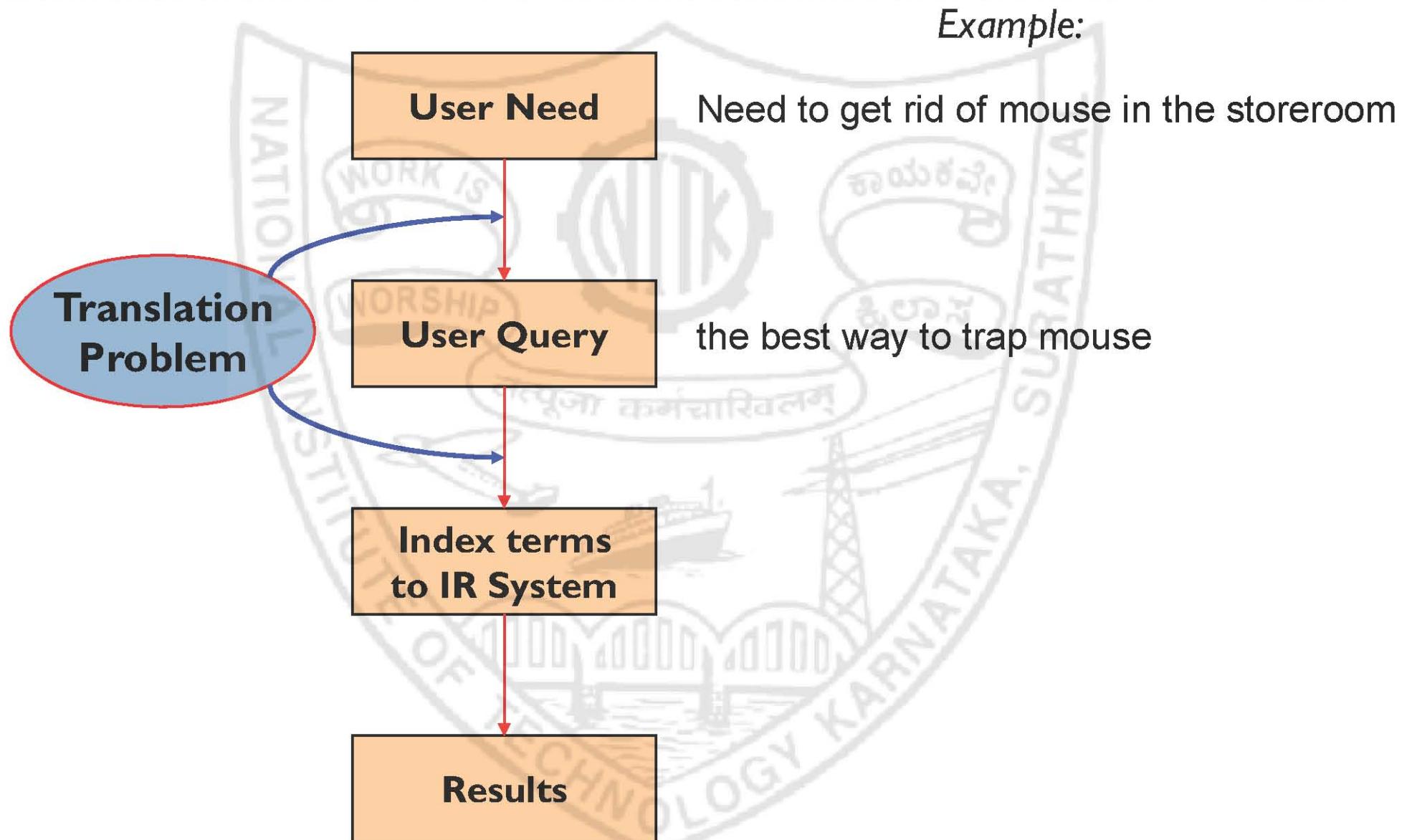
Why Don't Users Get What They Want?



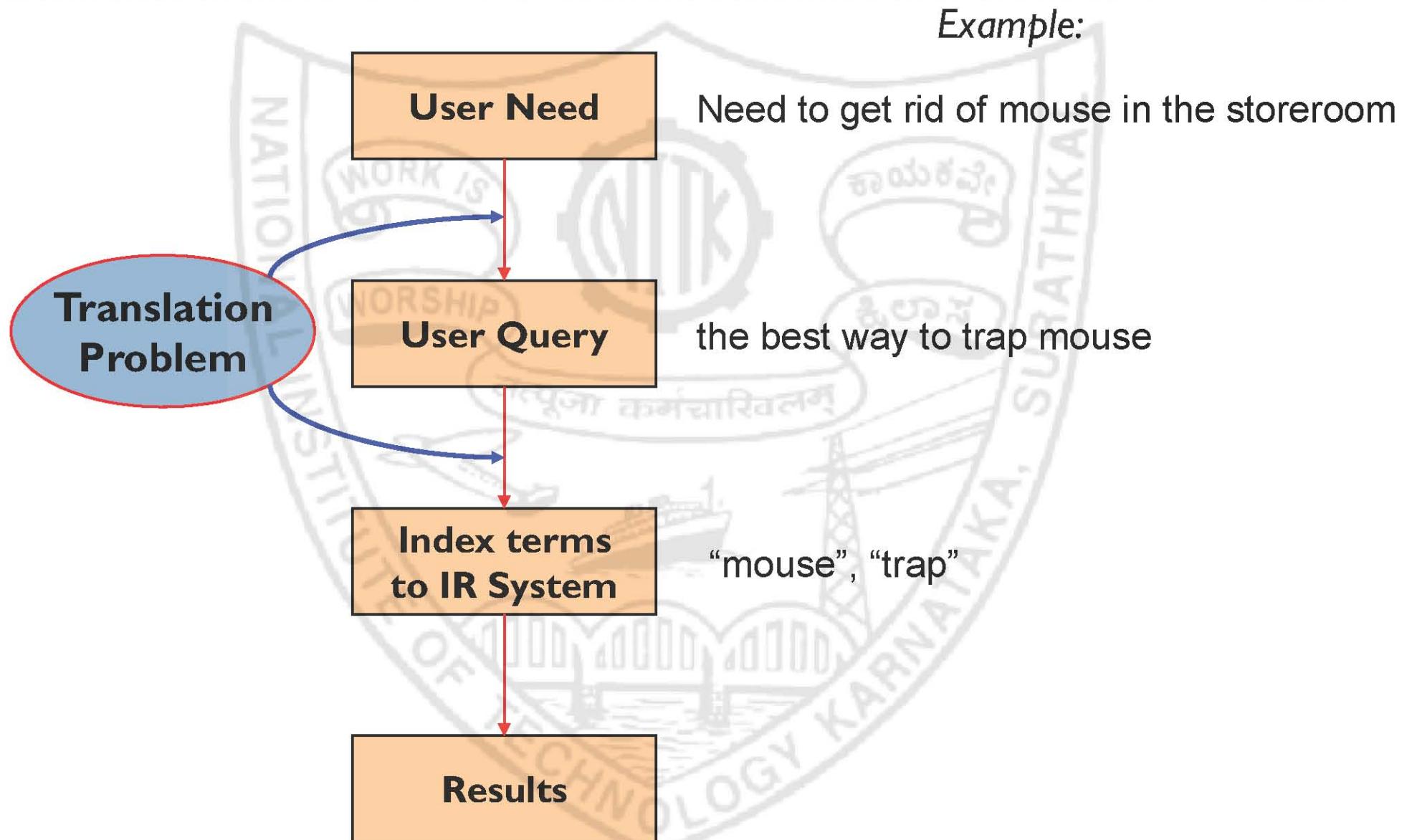
Why Don't Users Get What They Want?



Why Don't Users Get What They Want?



Why Don't Users Get What They Want?



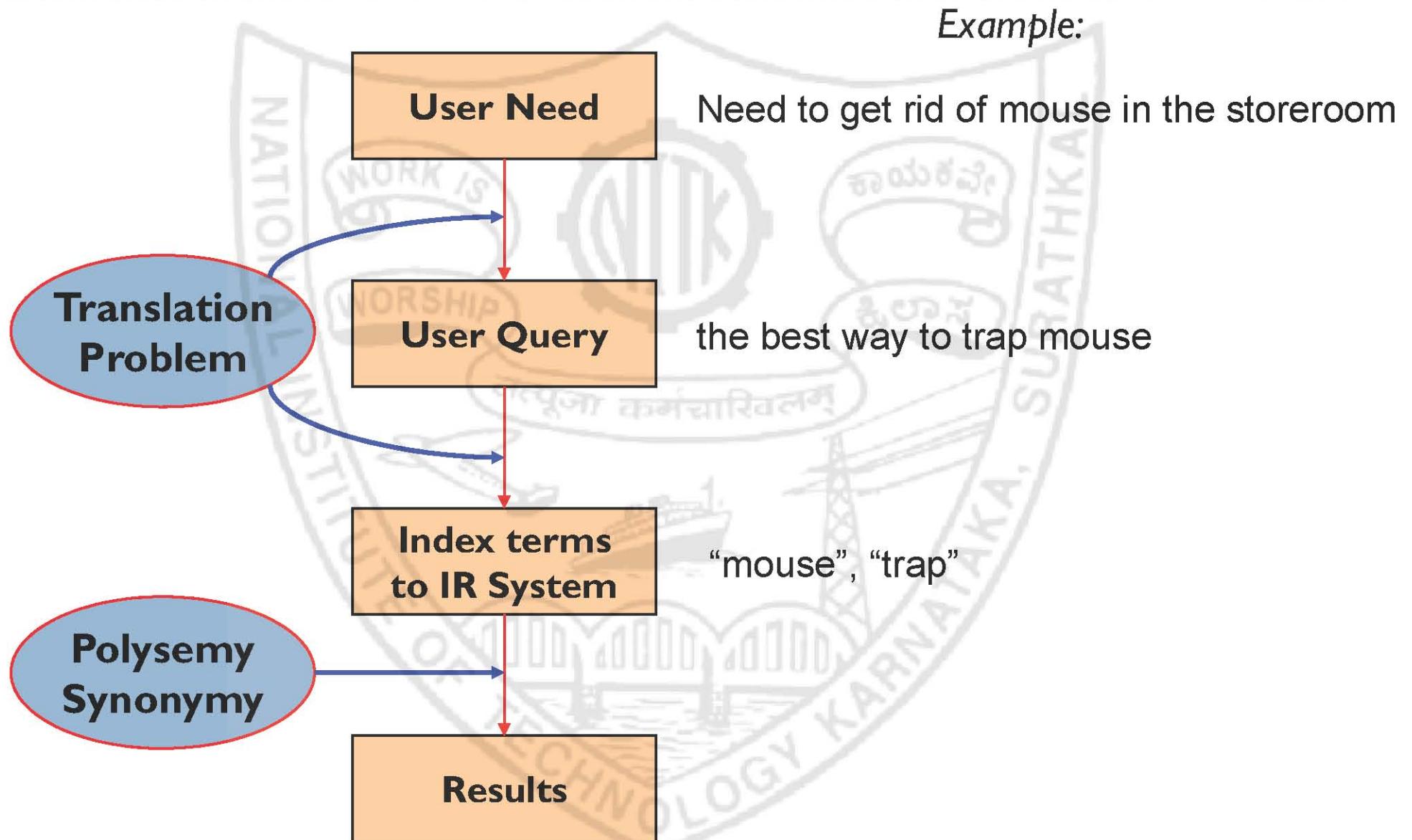
Example:

Need to get rid of mouse in the storeroom

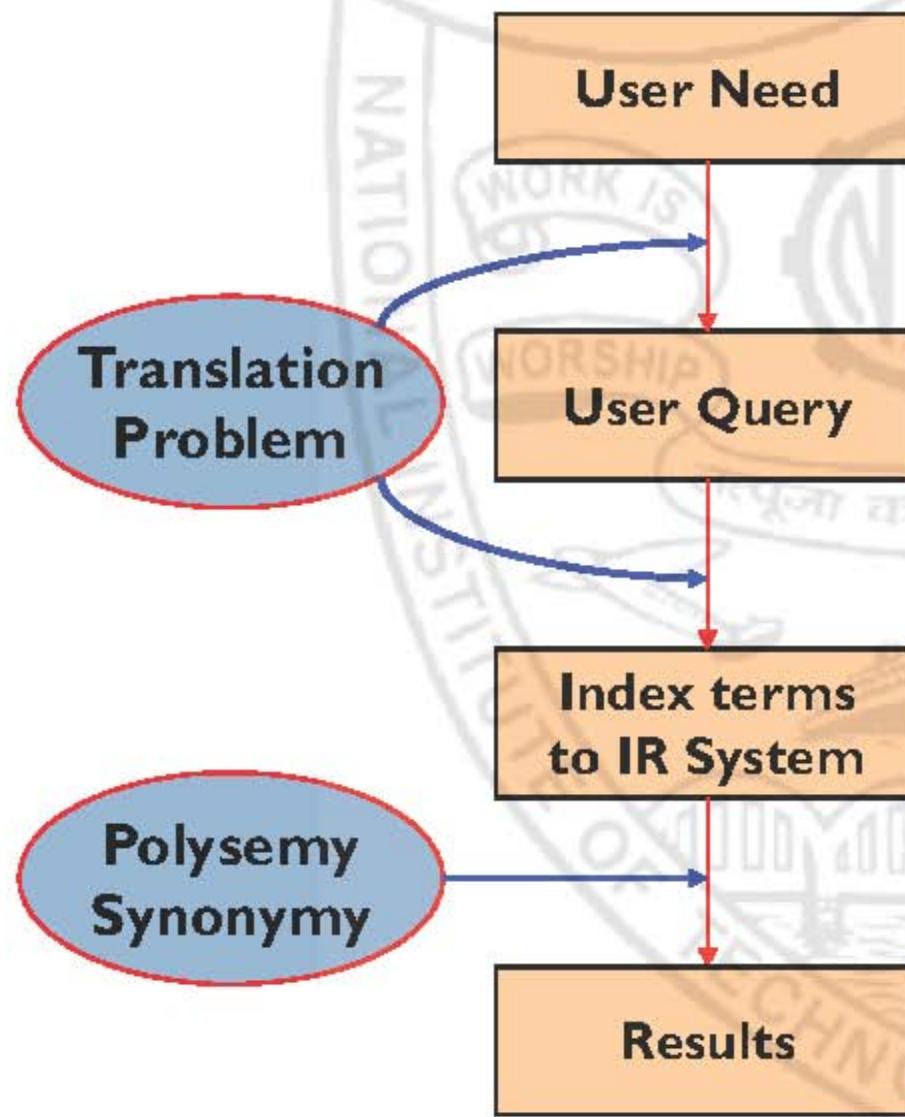
the best way to trap mouse

“mouse”, “trap”

Why Don't Users Get What They Want?



Why Don't Users Get What They Want?

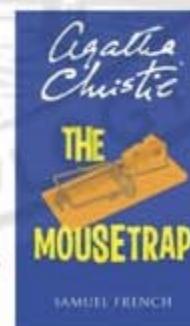


Example:

Need to get rid of mouse in the storeroom

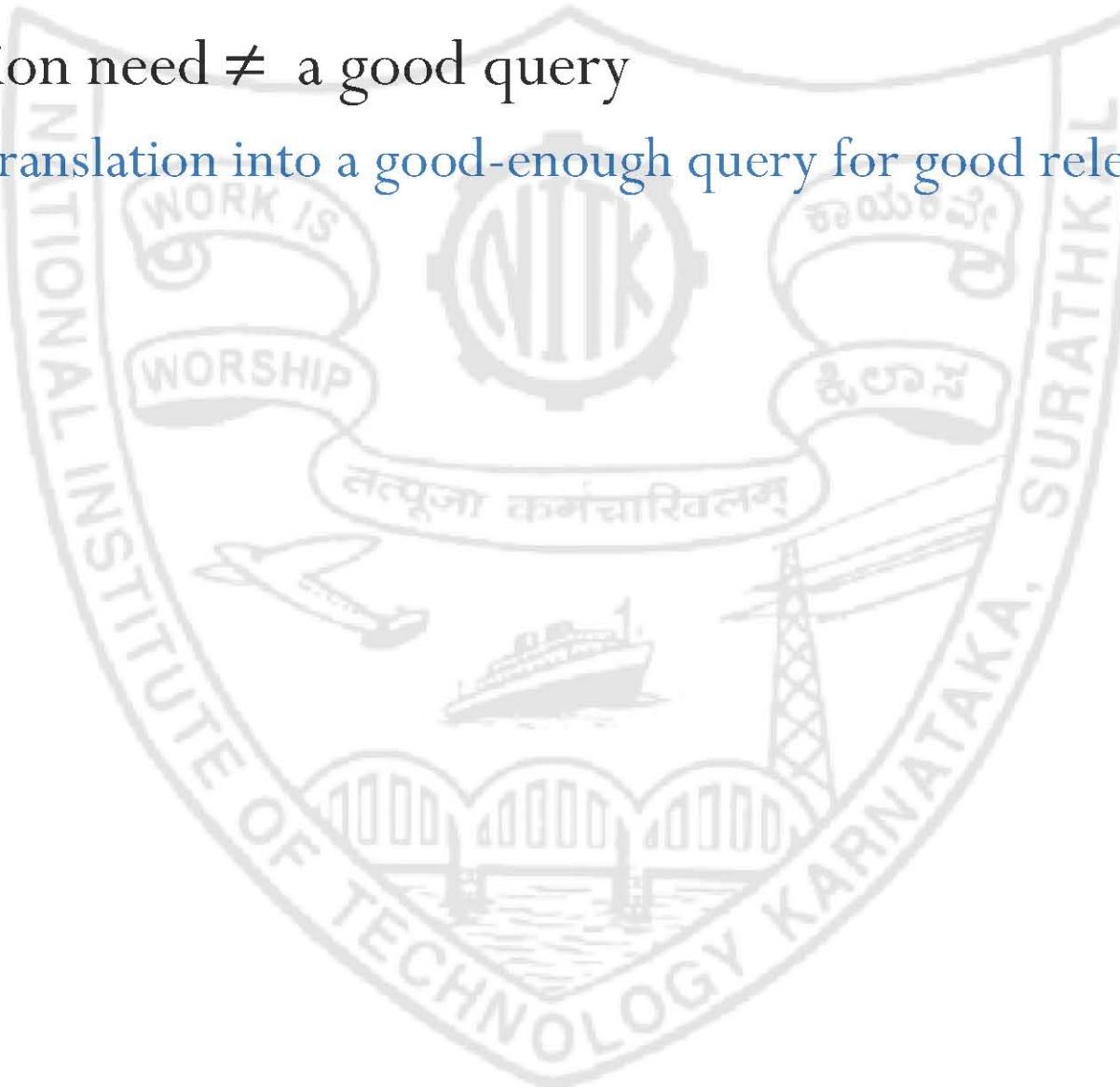
the best way to trap mouse

“mouse”, “trap”



The IR Problem

- ▶ Information need \neq a good query
 - ▶ Needs translation into a good-enough query for good relevancy.

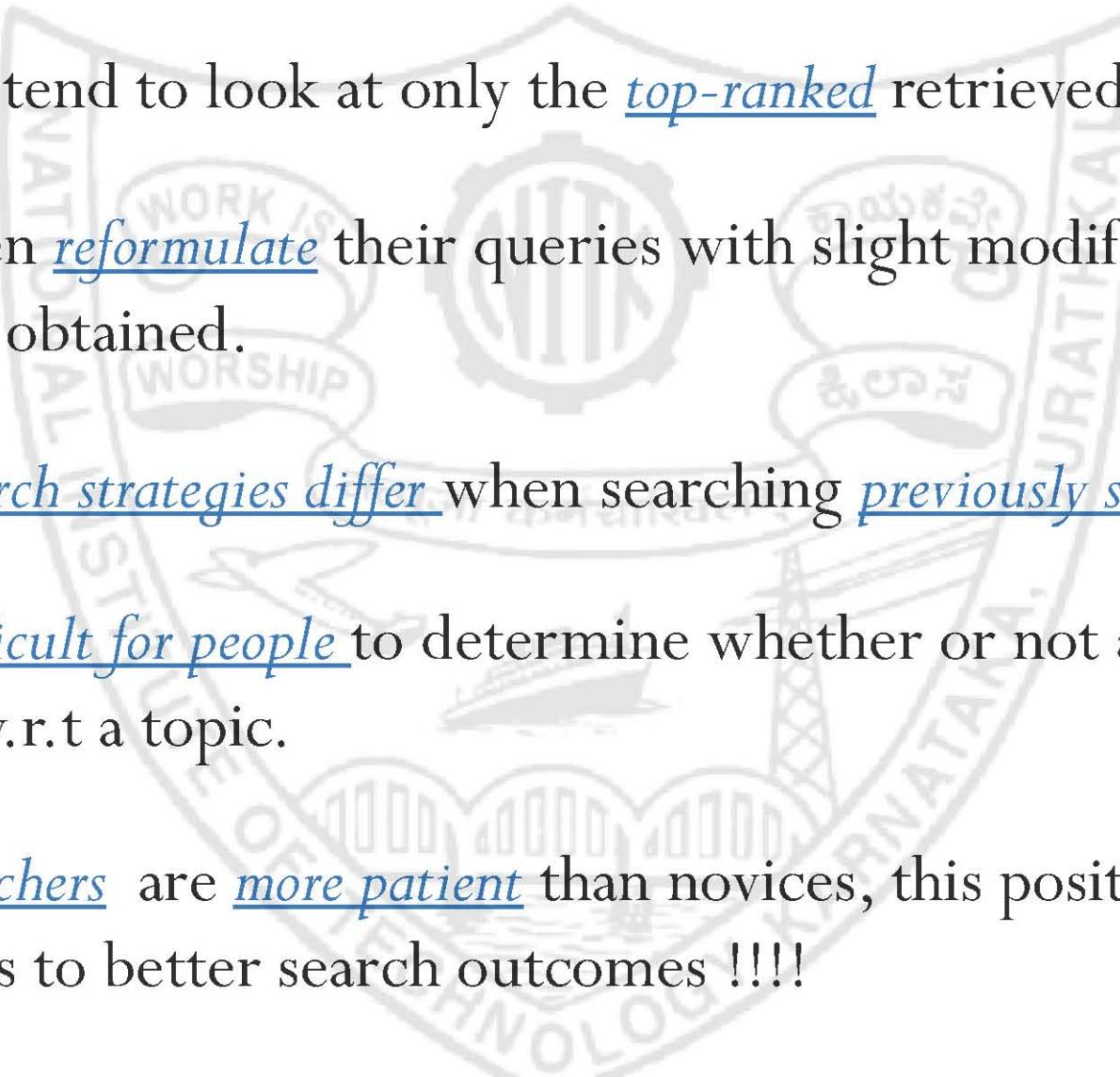


The IR Problem

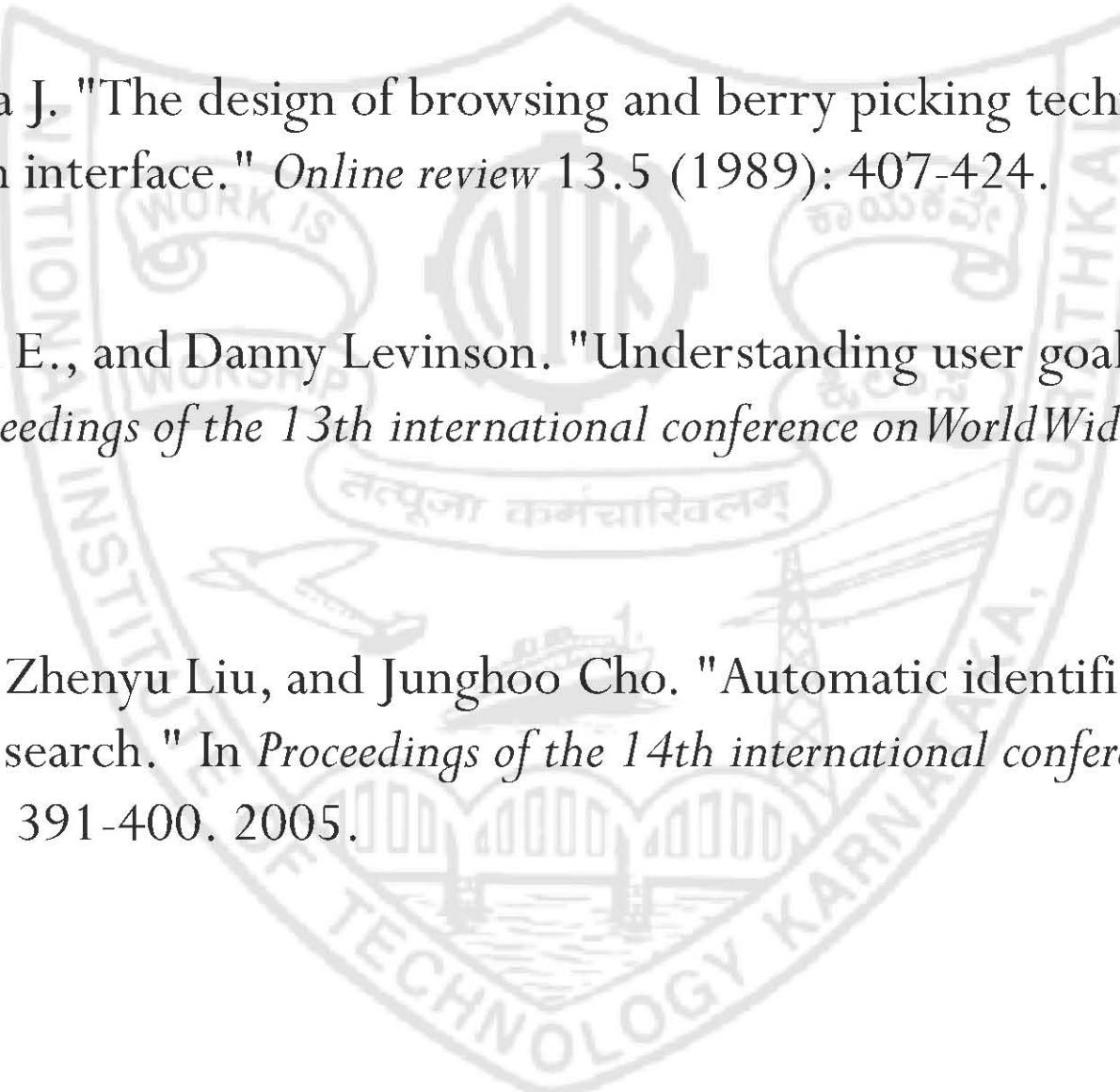
- ▶ Information need \neq a good query
 - ▶ Needs translation into a good enough query for good relevancy.
- ▶ The IR Problem -

“The key goal of an IR system is to retrieve all the items that are relevant to a user query, while retrieving as few non-relevant items as possible”

The Search Process – Summary

- 
- ▶ Searchers tend to look at only the top-ranked retrieved results.
 - ▶ Users often reformulate their queries with slight modifications, based on results obtained.
 - ▶ Users' search strategies differ when searching previously seen info.
 - ▶ Often difficult for people to determine whether or not a document relevance w.r.t a topic.
 - ▶ Expert searchers are more patient than novices, this positive attitude often leads to better search outcomes !!!!
-
- ▶ Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal
 - ▶ 3-Aug-22

More reading...

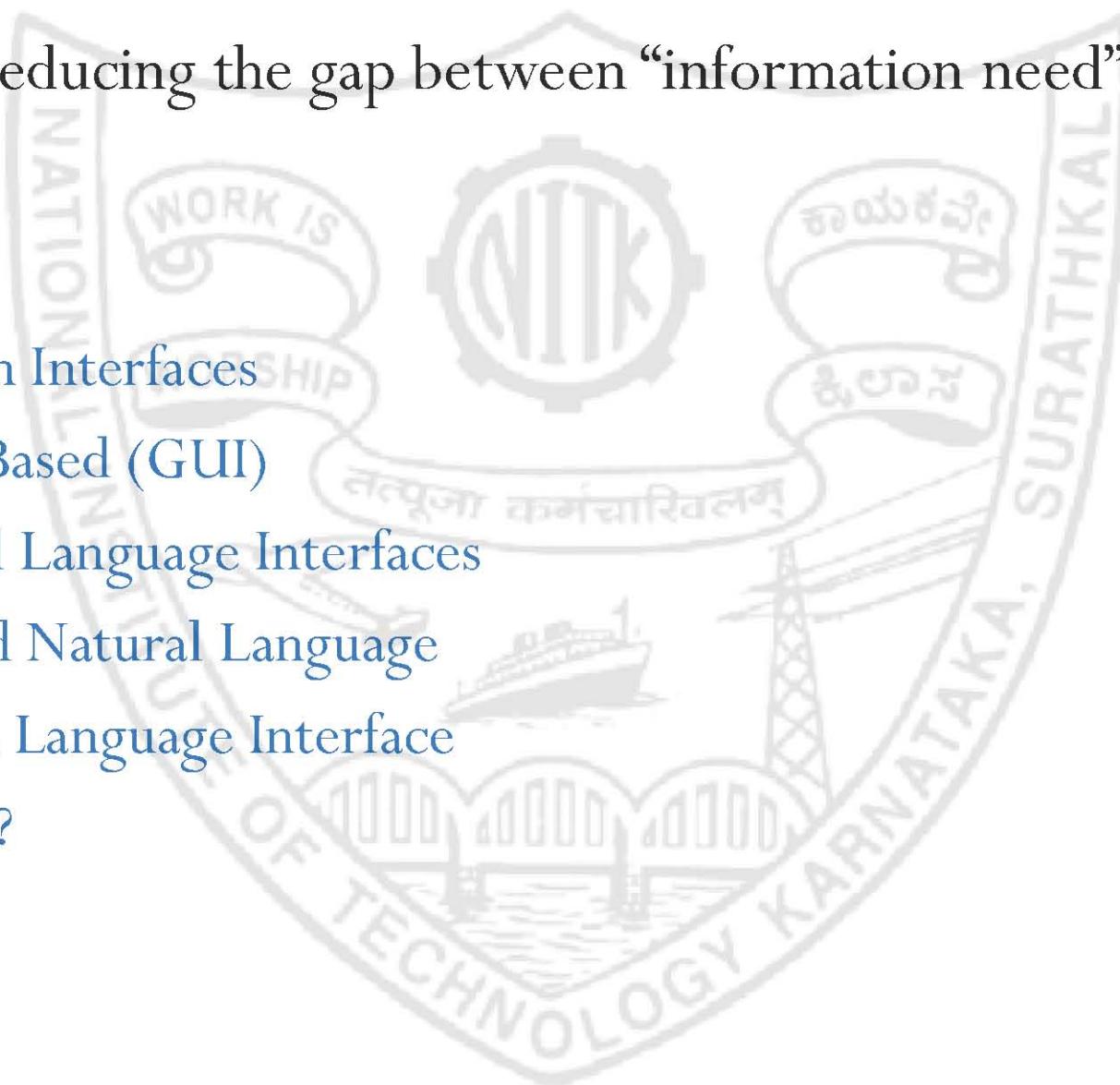
- 
- ▶ Bates, Marcia J. "The design of browsing and berry picking techniques for the online search interface." *Online review* 13.5 (1989): 407-424.
 - ▶ Rose, Daniel E., and Danny Levinson. "Understanding user goals in web search." *Proceedings of the 13th international conference on World Wide Web*. ACM, 2004.
 - ▶ Lee, Uichin, Zhenyu Liu, and Junghoo Cho. "Automatic identification of user goals in web search." In *Proceedings of the 14th international conference on World Wide Web*, pp. 391-400. 2005.
-
- ▶ Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

One possible Solution: Intuitive Search Interfaces



Modern Search Interfaces...

- ▶ Help in reducing the gap between “information need” and “query”.
- ▶ Types:
 - ▶ Boolean Interfaces
 - ▶ Form-Based (GUI)
 - ▶ Natural Language Interfaces
 - ▶ Stylized Natural Language
 - ▶ Spoken Language Interface
 - ▶ Others?



Modern Search Interfaces...

- ▶ Features of modern search interfaces:
 - ▶ auto-complete
 - ▶ auto-suggest
 - ▶ auto-correct
 - ▶ popular queries
 - ▶ dynamic query suggestions
 - ▶ ...

