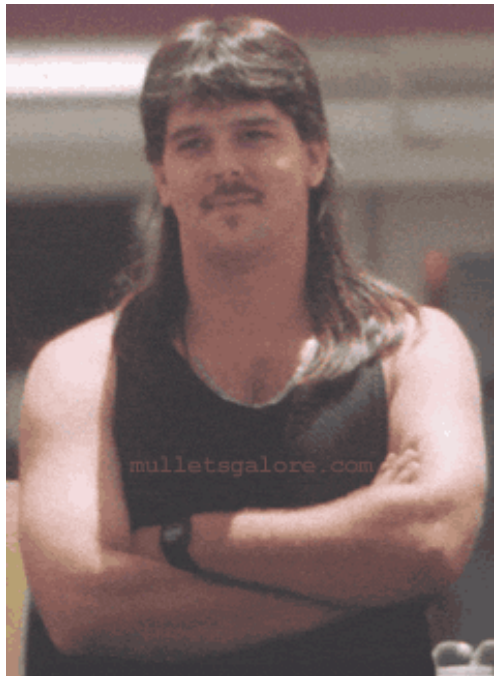
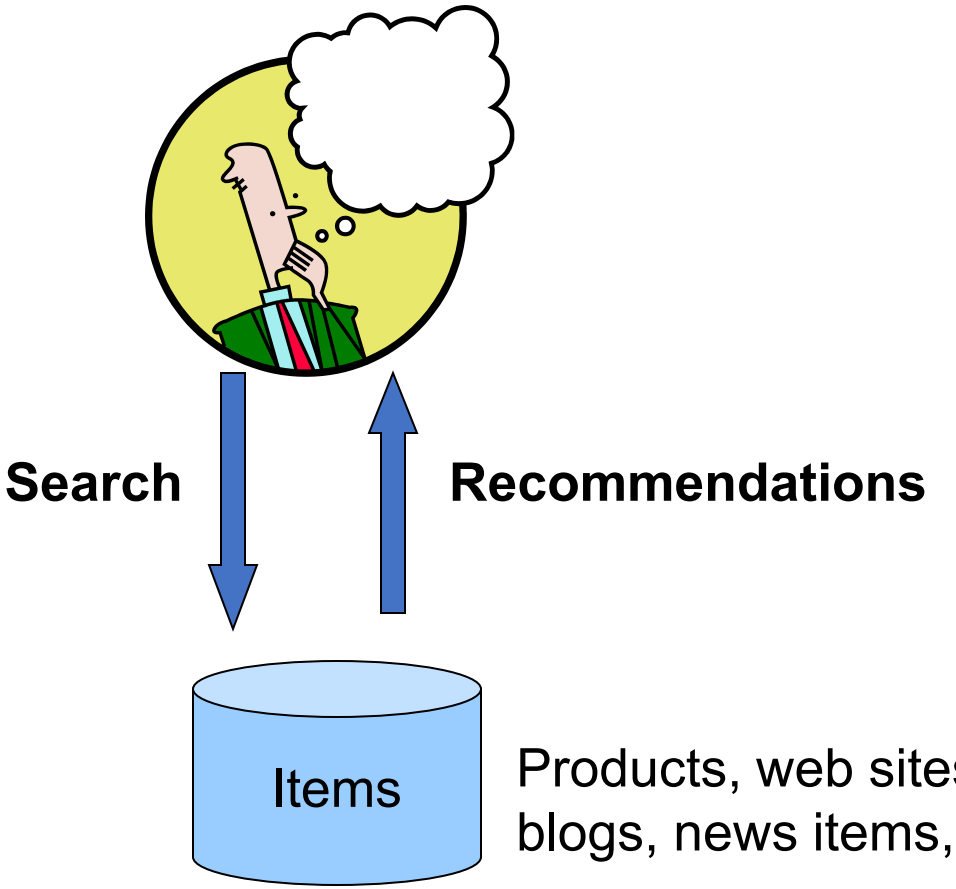
Recommender systems

Example: Recommender Systems



- **Customer X**
  - Buys Metallica CD
  - Buys Megadeth CD
- **Customer Y**
  - Does search on Metallica
  - Recommender system suggests Megadeth from data collected about **customer X**

Recommendations



Products, web sites, blogs, news items, …

Examples:

amazon.com
PANDORA
StumbleUpon
del.icio.us
NETFLIX
movielens
helping you find the *right* movies
last·fm
the social music revolution
Google News
YouTube
XBOX LIVE

Motivation for Recommender Systems

- Automates quotes like:
  - "I like this book; you might be interested in it"
  - "I saw this movie, you'll like it“
  - "Don't go see that movie!"

Further Motivation

Many of the top commerce sites use recommender systems to improve sales.

Users may find new books, music, or movies that was previously unknown to them.

Also can find the opposite for e.g.: movies or music that will definitely not be enjoyed.

Recommender System Types

Collaborative/Social-filtering system – aggregation of consumers' preferences and recommendations to other users based on similarity in behavioral patterns

Content-based system – supervised machine learning used to induce a classifier to discriminate between interesting and uninteresting items for the user

Content-based Recommendations

Main idea: Recommend items to customer ***x*** similar to previous items rated highly by ***x***

Example:

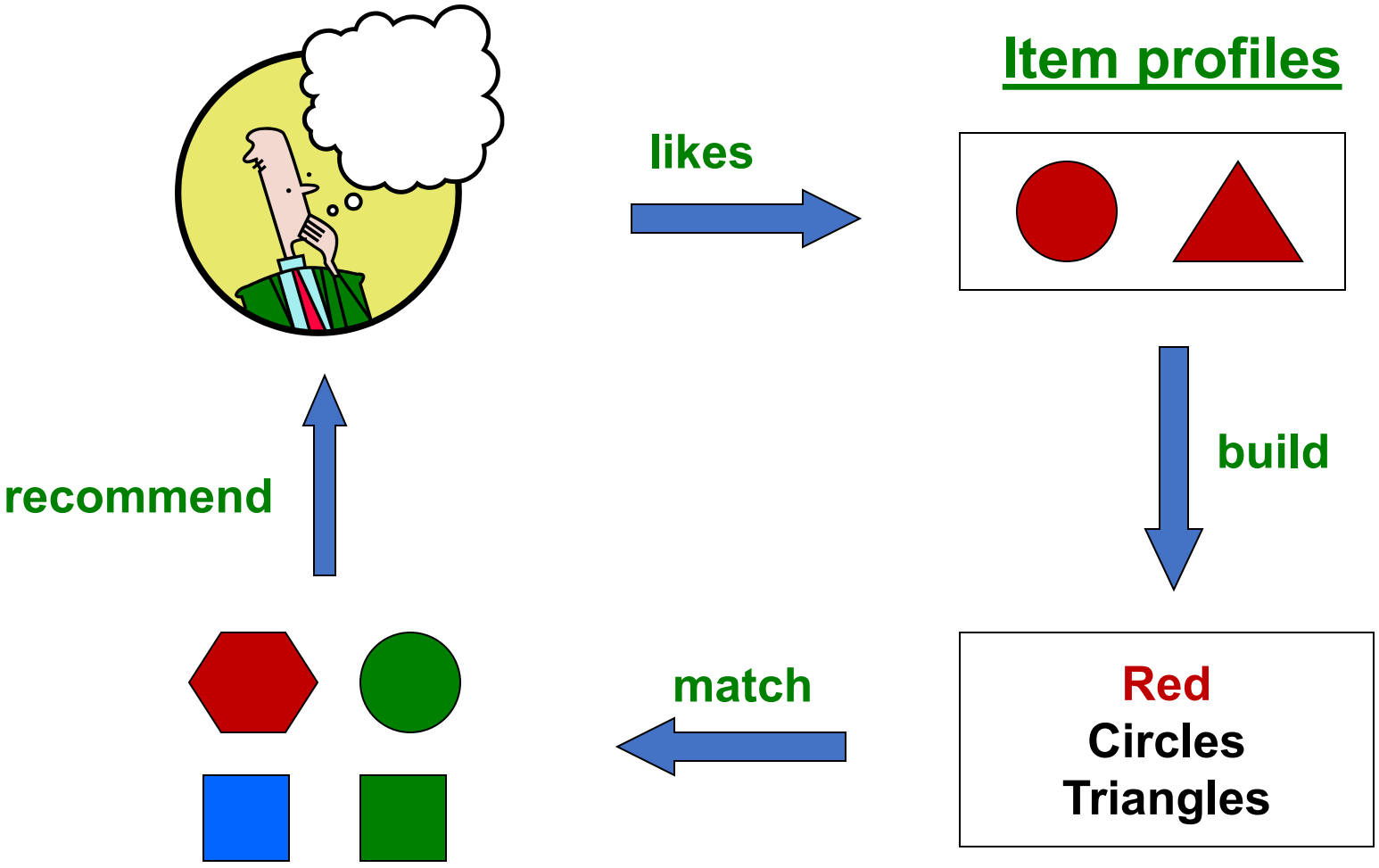
Movie recommendations

Recommend movies with same actor(s), director, genre, …

Websites, blogs, news

Recommend other sites with “similar” content

Plan of Action



Pros: Content-based Approach

- **+: No need for data on other users**
  - No cold-start or sparsity problems
- **+: Able to recommend to users with unique tastes**
- **+: Able to recommend new & unpopular items**
  - No first-rater problem
- **+: Able to provide explanations**
  - Can provide explanations of recommended items by listing content-features that caused an item to be recommended

Cons: Content-based Approach

- **–: Finding the appropriate features is hard**
  - E.g., images, movies, music
- **–: Recommendations for new users**
  - **How to build a user profile?**
- **–: Overspecialization**
  - Never recommends items outside user's content profile
  - People might have multiple interests
  - **Unable to exploit quality judgments of other users**

Collaborative filtering

Word of mouth

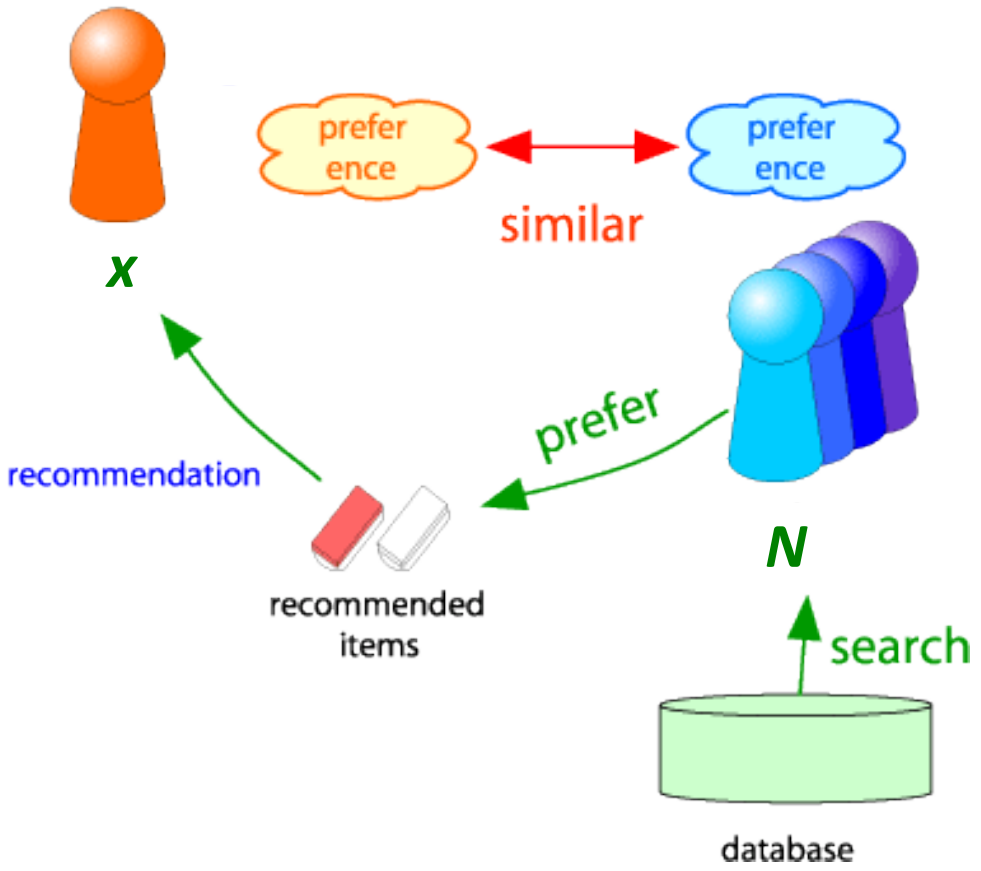
"The process in which the purchaser of a product or service tells friends, family, neighbors, and associates about its virtues, especially when this happens in advance of media advertising."

In practice I read book, recommend it to friends.

On e-commerce, Buy book-rate it-friend logs in-deduce similar taste- notice not bought book-recommend highly rated

Collaborative Filtering

- Consider user ***x***
- Find set ***N*** of other users whose ratings are "**similar**" to ***x***'s ratings
- Estimate ***x***'s ratings based on ratings of users in ***N***



Collaborative Filtering

	Star Wars	Hoop Dreams	Contact	Titanic
Joe	5	2	5	4
John	2	5		3
Al	2	2	4	2
Nathan	5	1	5	?

The problem of collaborative filtering is to predict how well a user will like an item that he has not rated given a set of historical preference judgments for a community of users.

Collaborative Filtering

User based CF

- Recommends items by finding similar users to the *active user* (to whom we are trying to recommend a movie).

Item based CF

- For item ***i***, find other similar items
- Estimate rating for item ***i*** based on ratings for similar items

Pros/Cons of Collaborative Filtering

- **+ Works for any kind of item**
  - No feature selection needed
- **- Cold Start:**
  - Need enough users in the system to find a match
- **- Sparsity:**
  - The user/ratings matrix is sparse
  - Hard to find users that have rated the same items
- **- First rater:**
  - Cannot recommend an item that has not been previously rated
  - New items, Esoteric items
- **- Popularity bias:**
  - Cannot recommend items to someone with unique taste
  - Tends to recommend popular items

CB V/S CF

Content-based Filtering	Collaborative Filtering
I. Its is based on concept “Show me more of what I have liked.”	i. This is basically people to people correlation.
II. It takes into account user preferences and on the basis of that makes the recommendations.	ii. It uses the wisdom of the crowd to recommend items to the user.
III. It recommends those items which are similar to user preferences based on their past behaviour.	iii. The ratings can either be explicit or implicit. It assumes that people who had similar tastes in past, will also have similar tastes in future also.

Snippet Generation

- Query-dependent document summary
- Simple summarization approach
  - rank each sentence in a document using a *significance factor*
  - select the top sentences for the summary
  - first proposed by Luhn in 50’s

Tropical Fish

One of the U.K.s Leading suppliers of **Tropical**, Coldwater, Marine **Fish** and Invertebrates plus.. . next day **fish** delivery service ...

www.**tropicalfish**.org.uk/**tropical_fish**.htm Cached page

Firewall Authentication Keepaliv | snippet generation in informatic | Information Retrieval | Levenshtein Distance | Levenshtein Distance | download

https://www.google.co.in/search?q=snippet+generation+in+information+retrieval+ppt&ei=s3dzXL_VCIzkvgTiwYSIBg&start=0&sa=N&ved=0ahUKEwj_-uSpjtbgAhUMso8KHelgAWE4ChDy0wMIfQ&biw=1600&bih=708

Apps | cursor | Create a Slogan for ... | eBook/Python 3 Ob... | How to use form 16... | Data Structures and... | MCQ Quizzes on D... | Data Structure Mult... | Byte-Python-concu... | Introduction to Pro... | Python Threads, pyt...

Google

snippet generation in information retrieval ppt

All | Images | News | Videos | Shopping | More | Settings | Tools

About 1,46,000 results (0.51 seconds)

Scholarly articles for **snippet generation in information retrieval ppt**

... cyberlearning resources through **information retrieval** ... - Liu - Cited by 19
Customizing **information** by combining pair of ... - Pierre - Cited by 50
Exploratory web searching with dynamic taxonomies ... - Papadakos - Cited by 26

[PPT] **Information Retrieval - College of Engineering and Computer Science**
cecs.wright.edu/~tkprasad/courses/cs707/L10Evaluation.ppt
"KWIC" **snippets**: Keyword in Context **presentation**; Generated in conjunction with scoring. If query found as ... **Generating** dynamic summaries. If we have only a ... But users really like **snippets**, even if they complicate **IR** system design. Prasad.

[PPT] **Information Retrieval (IR) - UCSB Computer Science - UC Santa Barbara**
www.cs.ucsb.edu/~tyang/class/290N15/slides/Topic1SearchIntroSimple.ppt
Introduction to **Information Retrieval** and Advanced Internet Services ... **generation** ... Constructs the display of ranked documents for a query; Generates **snippets** (dynamic description) to show how queries match documents; Highlights ...

[PDF] **Index-based Snippet Generation - CiteSeerX**
citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.583.2790&rep=rep1...
by G Manolache - 2008 - Cited by 5 - Related articles
Jun 9, 2008 - 5 Index-based versus Document-based **Snippet Generation** ... The usual way of interacting with an **IR** system is to enter a specific information need .... sentence as the minimal unit for extraction and **presentation** to the user.

[PDF] **6 Efficient Index-Based Snippet Generation - Documentation**

group4-QuestionA....ppt | 08eval.pptx | Show all

ENG INTL 10:44 25-02-2019

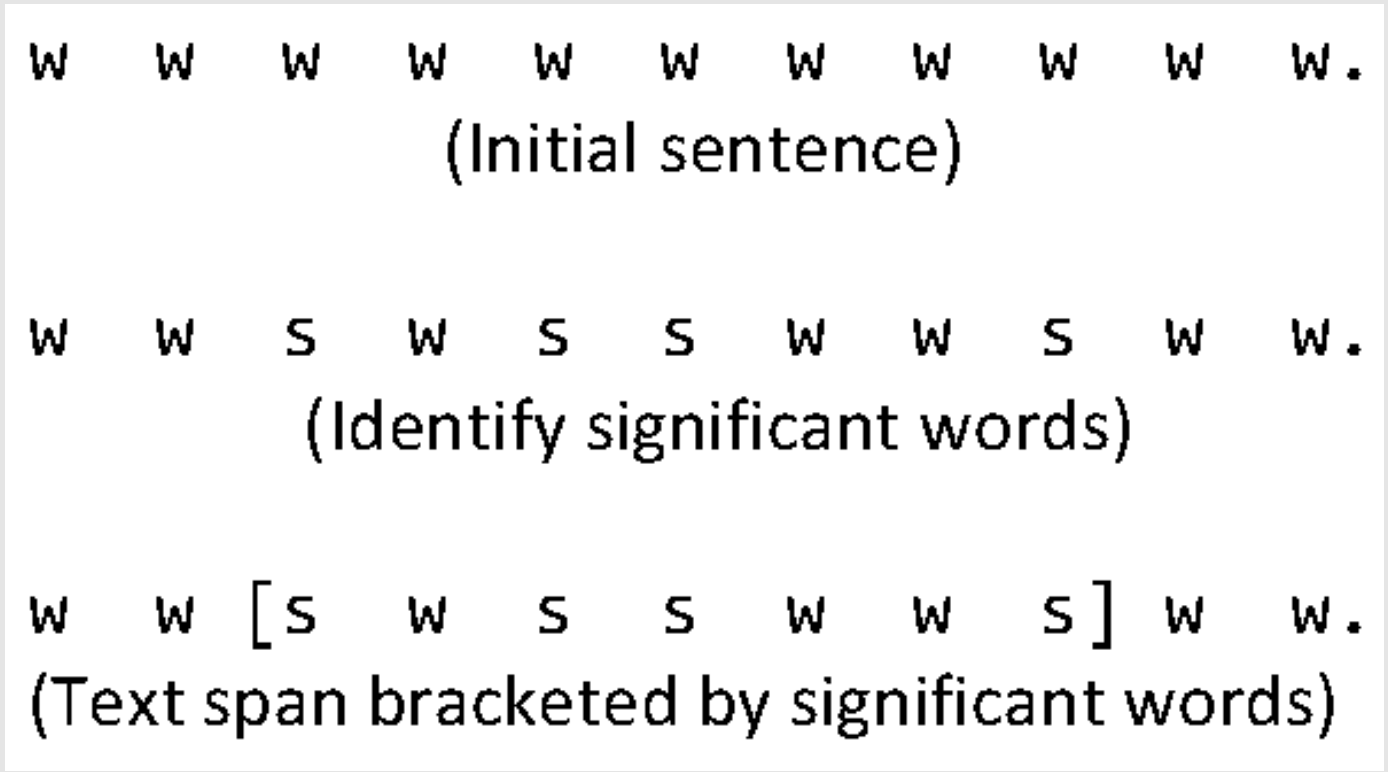
Sentence Selection

- Significance factor for a sentence is calculated based on the occurrence of *significant words*
  - If $f_{d,w}$ is the frequency of word *w* in document *d*, then *w* is a significant word if it is not a stopword and

$$f_{d,w} \geq \begin{cases} 7 - 0.1 \times (25 - s_d), & \text{if } s_d < 25 \\ 7, & \text{if } 25 \leq s_d \leq 40 \\ 7 + 0.1 \times (s_d - 40), & \text{otherwise} \end{cases}$$

    where s_d is the number of sentences in document *d*
  - text is *bracketed* by significant words (limit on number of non-significant words in bracket)

Sentence Selection



- Significance factor for bracketed text spans is computed by dividing the square of the number of significant words in the span by the total number of words
- e.g.,
- Significance factor = $4^2/7$ = 2.3

Snippet Generation

- Involves more features than just significance factor
- e.g. for a news story, could use
  - whether the sentence is a heading
  - whether it is the first or second line of the document
  - the total number of query terms occurring in the sentence
  - the number of unique query terms in the sentence
  - the longest contiguous run of query words in the sentence
  - a density measure of query words (significance factor)
- Weighted combination of features used to rank sentences

Snippet Generation

- Web pages are less structured than news stories
  - can be difficult to find good summary sentences
- Snippet sentences are often selected from other sources
  - metadata associated with the web page
    - e.g., <meta name="description" content= ...>
  - external sources such as web directories
    - e.g., Open Directory Project, http://www.dmoz.org
- Snippets can be generated from text of pages like Wikipedia

Snippet Guidelines

All query terms should appear in the summary, showing their relationship to the retrieved page

When query terms are present in the title, they need not be repeated

allows snippets that do not contain query terms

Highlight query terms in URLs

Snippets should be readable text, not lists of keywords

What is summarization?

A summary is a text that is produced from one or more texts, that contains a significant portion of the information in the original text(s), and that is no longer than half of the original text(s).

Summaries may be classified as:

Extractive	Abstractive

Extractive summaries

Extractive summaries are created by reusing portions (words, sentences, etc.) of the input text verbatim.

For example, search engines typically generate extractive summaries from webpages.

Most of the summarization research today is on extractive summarization.

Text:

Indian Institute of Technology Bombay

From Wikipedia, the free encyclopedia

This article **needs additional citations for verification**. Please h
adding citations to reliable sources. Unsourced material may be c
(September 2010)

The **Indian Institute of Technology Bombay** (abbreviated **IITB** or **IIT Bombay**) is a public engineering institution located in Powai, Mumbai, India. It has been ranked among the top engineering colleges in India.[1] It is the second-oldest institute of the Indian Institutes of Technology system.[2]

Extractive summary:

Indian Institute of Technology Bombay - Wikipedia, the free ...
en.wikipedia.org/wiki/Indian_Institute_of_Technology_Bombay
The **Indian Institute of Technology Bombay** (abbreviated **IITB** or **IIT Bombay**) is a public engineering institution located in Powai, Mumbai, India. It has been ...

Abstractive summaries

In abstractive summarization, information from the source text is re-phrased.

Human beings generally write abstractive summaries (except when they do their assignments ☺).

Abstractive summarization has not reached a mature stage because allied problems such as semantic representation, inference and natural language generation are relatively harder.

Abstractive summary: Book review

- An innocent hobbit of The Shire journeys with eight companions to the fires of Mount Doom to destroy the One Ring and the dark lord Sauron forever.

What is Question Answering?

Type of information retrieval. Given a collection of documents ,the system should be able to retrieve answers to questions posed in natural language.

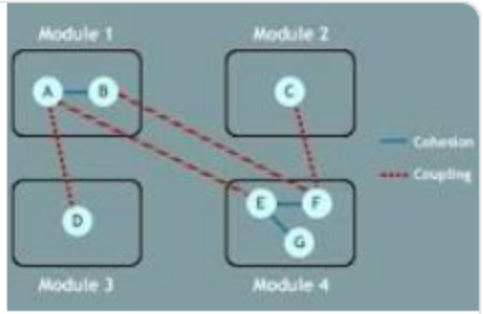
-Wikipedia

what is cohesion and coupling

All Videos Images Shopping News More Settings Tools

About 23,10,000 results (0.59 seconds)

Cohesion is a degree (quality) to which a component / module focuses on the single thing. **Coupling** is a degree to which a component / module is connected to the other modules. Aug 27, 2013



Difference between Cohesion and Coupling | Cohesion vs ...

https://freefeast.info › difference-between › difference-between-cohesion-an...

About Featured Snippets Feedback

People also ask

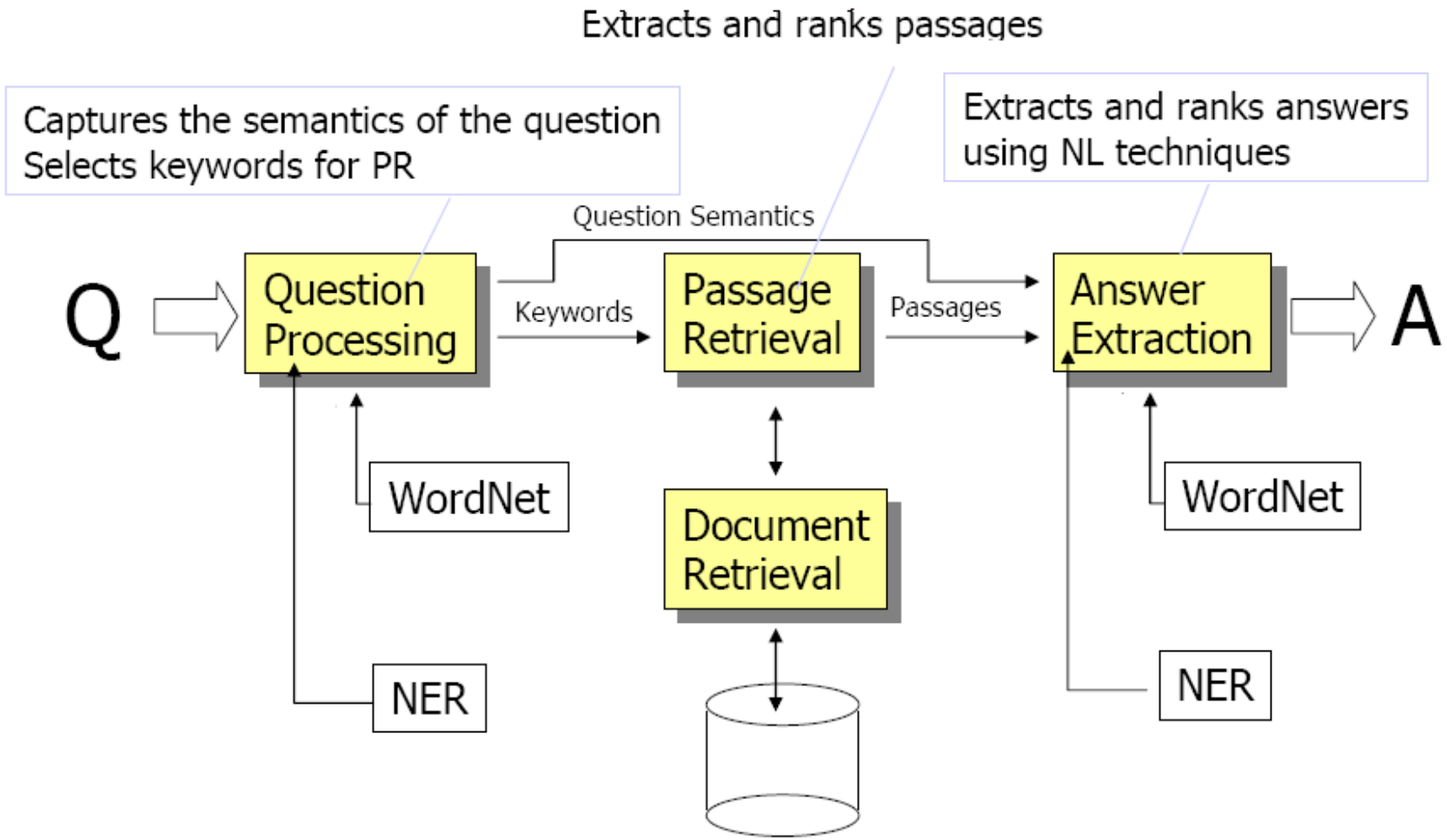
What is the difference between cohesion and coupling?

What is coupling and cohesion in system analysis and design?

What is cohesion in programming?

What is high cohesion and low coupling?

Generic QA Architecture



Question Answering

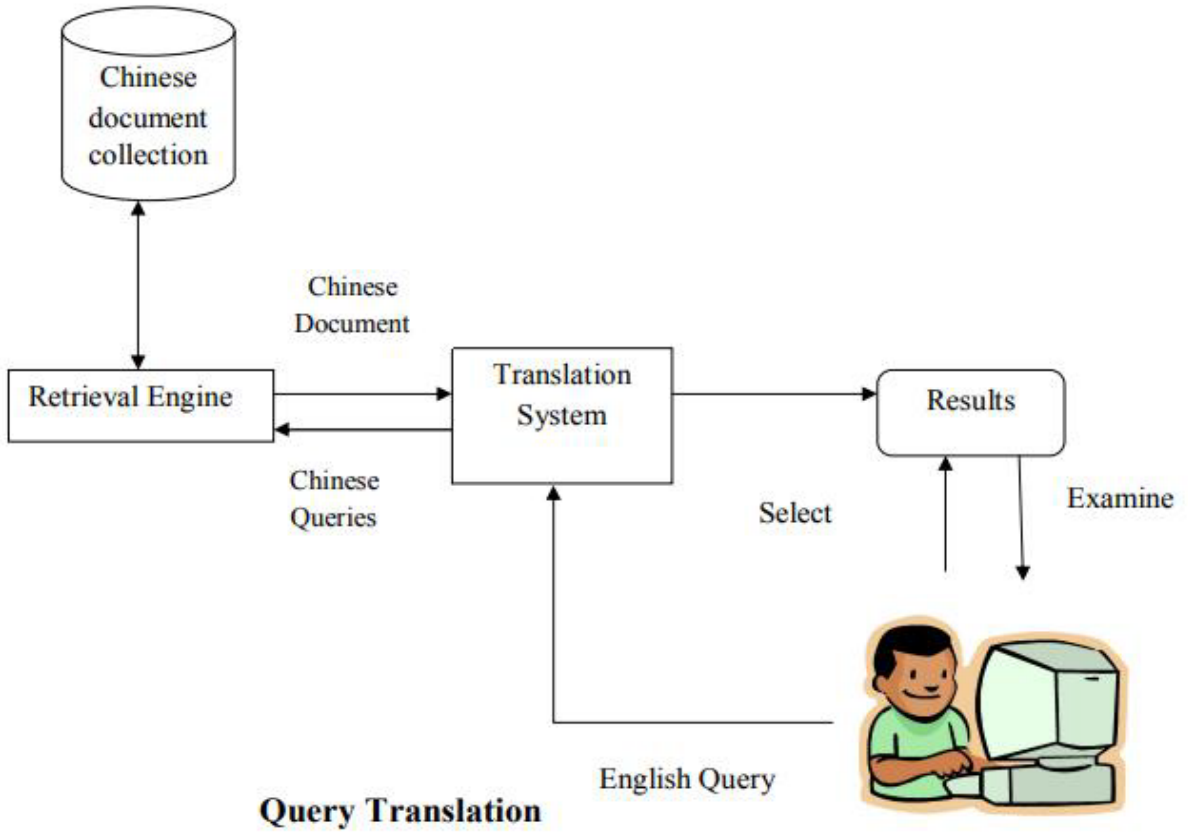
- QA systems can pull answers from an unstructured collection of natural language documents.
- QA research attempts to deal with a wide range of question types including: fact, list, definition, How, Why, hypothetical, semantically constrained, and cross-lingual questions.

CROSS LINGUAL

- Refers to the retrieval of documents that are in a language different from the one in which the query is expressed.
- Two methods : Query/Document Translation

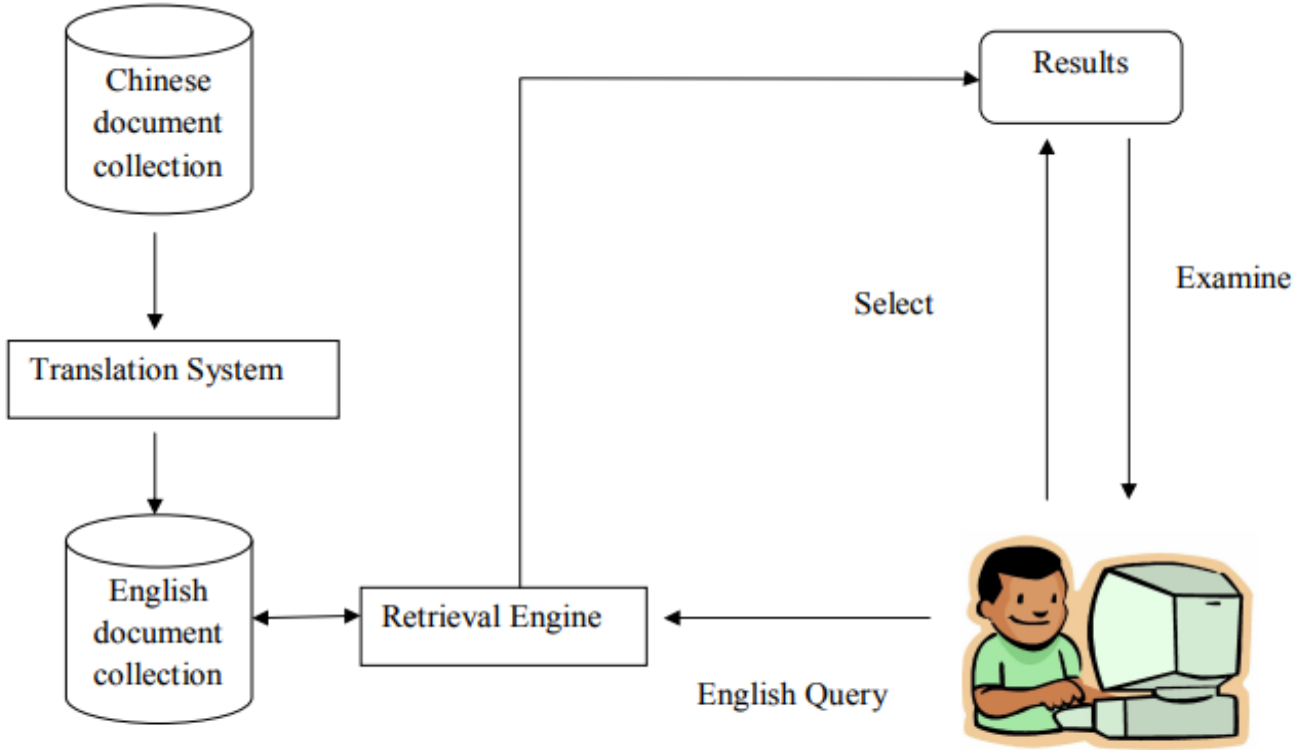
Query Translation

- Translate English query into Hindi query
- Search Hindi document collection
- Translate retrieved results back into English
- Query translation is easy.
- Translation of documents must be performed at query time.



Document translation

- Translate entire document collection into English
- Search collection in English
- Documents can be translated and stored offline. Automatic translation can be slow



Document Translation

Hidden/Invisible web

- The Invisible Web is the part of the World Wide Web, which is not indexable by search engines and is therefore invisible.
- In contrast to the Surface Web, the Invisible Web consists of data and information that cannot be searched with search engines for various reasons.
- Users cannot access this information by using traditional search engines.
- Non-indexed websites, apps, and resources include protected information in the areas of email, online banking, specialized databases, and other paid services