Efficient Information Retrieval Using Measures of Semantic Similarity

Krishna Sapkota
Center for Research in Social Defense Technology,
Language Technology Group, *nec*
Student Affiliate

Laxman Thapa
Nepal Engineering College (*nec*)

Shailesh Pandey
Center for Research in Social Defense Technology,
Language Technology Group, *nec*
Research Engineer

Introduction

- Information in WWW are scattered and diverse in nature
- Users frequently fails to describe the information to retrieve
- Traditional search techniques are constrained by keyword based matching techniques
  - Hence low precision and recall is obtained

Limitations of Traditional Search

- Miss to retrieve synonymy terms
- Users must be intelligent
- Do not retrieve conceptual terms

Hence there is the need for semantic feature in search

Foundations

- Semantic similarity and relatedness
  - Degree of closeness of meaning of words
- Semantic Similarity
  - For example car and bicycle are more similar than car and human
- Semantic Relatedness
  - Car and bicycle are similar but hot and cold are more related
- Similarity applies to Noun but relatedness covers all category

Approaches for Computing Semantic Similarity

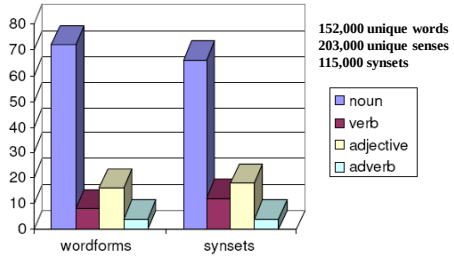
- Dictionary based
- Thesaurus based
- Semantic Network
  - Using path or node
  - Using Information Content

Word Pairs	Humans	Sim	Sim_{lin}	Sim'_{jcn}
fruit-furnace	0.05	1.85	0.14	0.05
monk-slave	0.57	2.53	0.21	0.05
coast-hill	1.26	6.19	0.53	0.09
magician-oracle	1.82	13.5	0.96	1.00
brother-lad	2.41	2.53	0.23	0.06
food-fruit	2.69	1.50	0.22	0.09
furnace-stove	3.11	1.85	0.13	0.04
boy-lad	3.82	8.29	0.72	0.18
automobile-car	3.92	8.62	1.00	maximum

WordNet Semantic Network

- Developed at the cognitive science laboratory of Princeton University
- English online lexical database
- Based on psycholinguistic principle
- Represents World Knowledge
- Four syntactic categories noun, verb, adverb, adjective are organized in to different relationships

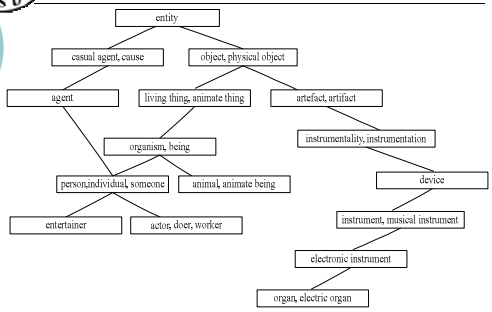
WordNet 2.0 Statistics



WordNet Relations

- Semantic relations
  - Hypernymy/hyponymy or IS-A
  - Meronymy/holonymy or HAS-A
  - Entailment
- Lexical relation
  - Synonymy
  - Antonymy

WordNet Hypernyms



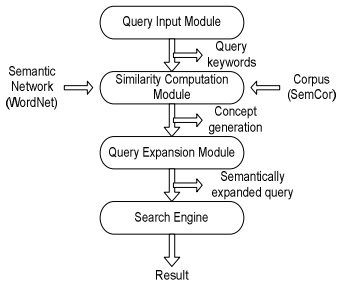
Extending the Existing Search Engine

- Low cost & less time rather than to built from scratch
- Our system uses Google API to perform search
- Google API can be accessed using SOAP from different programming language
- The users original queries are expanded and fed to Google Search Engine
  - Hence our system lies in between users and Google's interface

Our Algorithm

- Original query as "Bank Interest"
  - Bank has 10 concept and interest has 7 concept in WordNet 2.0
- To get optimized query
  - calculate similarity score of each pair of concepts as 10x7=70 similarity score is obtained
  - Pick up the concept with highest value of similarity score
  - Replace the original query with synonyms and hypernyms of concept that has highest similarity score

System Architecture



Search Results

Topic Retrieved	WWW Address
The National Neopian Bank	http://www.neopets.com/bank.phtml
Bankrate.com	http://www.bankrate.com
Certificate of deposit of interest rates: Compare the best rate	http://www.bankrate.com/brm/rate/deposits
Bank Interest Calculator	http://www.digita.com/tisali/calculators/bankinterestcalculate/
Personal Banking System, savings, bank interest rate, tax	http://www.thisismoney.co.uk/saving
Indian Bank-Interest rates	http://www.indian-bank.com/interest.htm
National Australian Bank	http://www.national.com.au/business-solution/02253300.htm
Bank Interest	http://www.ato.gov.au/content/48327.htm
Infochoice Banking	http://www.infochoice.com.au/banking/default.asp
National Bank Interest Rate Graph	http://www.nbnz.co.nz/economics/interest/

First 10 results from Google search for 'bank interest'

Search Results Contd.

Topic Retrieved	WWW Address
Scholarly articles (for all expanded term)	http://www.google.com
Bloomberg.com: Financial Glossary	http://www.bloomberg.com/invest/glossary/bfglost.hym
FCAC-Glossary	www.fcac-acfc.gc.ca/eng/glossary.asp
Guide to organize a new state bank in Florida	http://www.flofr.com/banking/howtoorg.htm
PSI- Performance Solution International	http://www.goto-psi.com/glossary.htm
[pdf] How should financial institution and market should be structured	http://www.iadb.org/res/publications/pubfiles/
Women's wallstreet.com – glossary	http://www.wimenswallstreet.com/tools-resources/glossary/f.htm
Operational risk poses challenges to financial institution	http://knowledge.wharton.upenn.edu/article.cfm?articleid=582
FDIC:FDIC Banking Review	www.fdic.gov/bank/analytical/banking/2004nov/article1/index
BKD ,LLP- Financial Services	http://www.bkd.com/industry/financial-services

First 10 results from our extended Google for bank interest

Programming Languages and Tools

- Used PHP as it provides easy access to internet service
- Used MySQL format of WordNet 2.0
- All 186 files of semantic concordance were converted to MySQL format
- Used nuSOAP a implementation of SOAP architecture in PHP by nuSphere corporation
- Used Google API service

Further Enhancements

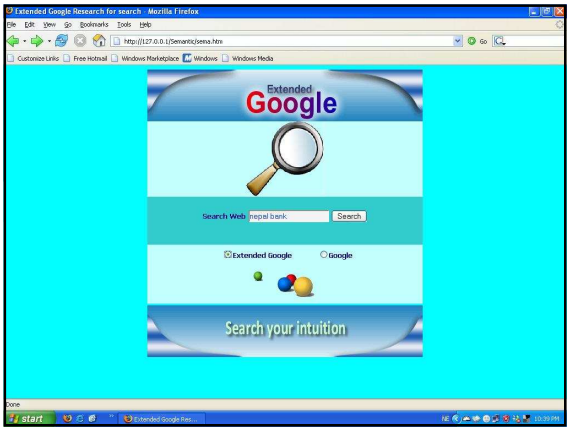
- Word Sense Disambiguation (WSD)
  - Process of assigning correct sense to the word
    - The chair for principal was not occupied. (position)
    - This chair is made up of good wood. (furniture)
- Measures for more syntactic category
- Customizable interface to search

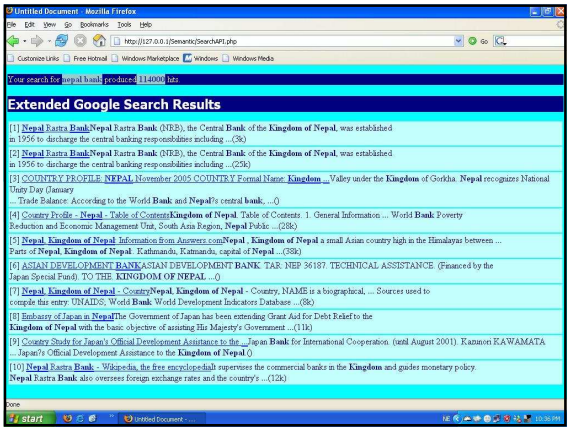
Conclusion

- Users queries were replaced by our analyzed terms
- User's must not be intelligent
- We provide Concept based search
- For better precision and recall incorporate WSD

THANK YOU

QUESTIONS?


Extended Google Research for search - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://127.0.0.1/Semantic/seena.htm
Customize Links Free Hotmail Windows Marketplace Windows Windows Media
Extended
Google
Search Web
nepal bank
Search
Extended Google
Google
Search your intuition
Done
start


Untitled Document - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://127.0.0.1/Semantic/SearchAPI.php
Customize Links Free Hotmail Windows Marketplace Windows Windows Media
Your search for nepal bank produced 114000 hits
Extended Google Search Results
[1] Nepal Rastra Bank:Nepal Rastra Bank (NRB), the Central Bank of the Kingdom of Nepal, was estabirhed in 1956 to dircharge the central banking responsibilies including ...(3k)
[2] Nepal Rastra Bank:Nepal Rastra Bank (NRB), the Central Bank of the Kingdom of Nepal, was estabirhed in 1956 to dircharge the central banking responsibilies including ...(23k)
[3] COUNTRY PROFILE: NEPAL November 2005 COUNTRY Formal Name: Kingdom ... Valley under the Kingdom of Gorkha. Nepal recognizes National Unity Day (January ... Trade Balance: According to the World Bank and Nepal?s central bank, ...()
[4] Country Profile - Nepal - Table of ContentsKingdom of Nepal. Table of Contents. 1. General Information ... World Bank Poverty Reduction and Economic Management Unit, South Asia Region, Nepal Public ...(28k)
[5] Nepal, Kingdom of Nepal: Information from Answers.comNepal , Kingdom of Nepal a small Asian country high in the Himalayas between ... Parts of Nepal, Kingdom of Nepal. Kathmandu, Katmandu, capital of Nepal ...(33k)
[6] ASIAN DEVELOPMENT BANKASIAN DEVELOPMENT BANK. TAR: NEP 36187. TECHNICAL ASSISTANCE. (Financed by the Japan Special Fund). TO THE KINGDOM OF NEPAL ...()
[7] Nepal, Kingdom of Nepal - CountryNepal, Kingdom of Nepal - Country, NAME is a biographical, ... Sources used to compile this entry: UNAID©, World Bank World Development Indicators Database ...(8k)
[8] Embassy of Japan in NepalThe Government of Japan has been extending Grant Aid for Debt Relief to the Kingdom of Nepal with the basic objective of assisting His Majesty's Government ...(11k)
[9] Country Study for Japan's Official Development Assistance to the ...Japan Bank for International Cooperation. (until August 2001). Kazunori KAWAMATA ... Japan?s Official Development Assistance to the Kingdom of Nepal ()
[10] Nepal Rastra Bank - Wikipedia, the free encyclopediaIt supervises the commercial banks in the Kingdom and guides monetary policy. Nepal Rastra Bank also oversees foreign exchange rates and the country's ...(12k)
Done
start