



Jan Pedersen
22 July 2010

Outline

- Problem Statement
 - Best effort retrieval vs automated reformulation
- Query Evaluation Architecture
- Query Understanding Models
 - Data Sources

Standard IR Assumptions

- Queries are well-formed expressions of intent
 - Best effort response to the query as given
- Users will reformulate
 - If results do not meet information need
- Actually, queries often contain errors
 - Misspelling
 - Incorrect terms
- Users can't reformulate if they don't understand what's wrong
 - Miss good content
 - Dead ends

Problem Definitions

- Best Effort Retrieval
 - Find the most relevant results for the user query
 - Segmentation
 - Stemming/Synonym expansion
 - Term deletion
- Automated Query Reformulation
 - Modify the user query to produce more relevant results for the inferred intent
 - Spell correction
 - Term deletion

Spelling Correction

The screenshot shows a Bing search interface. At the top left is the Bing logo. The search bar contains the text "blare house" and a magnifying glass icon. Below the search bar are tabs for "Web", "Local Listings", and "Wikipedia". A message reads: "We're including results for **blair house**. Do you want results only for **blare house**?" A blue callout bubble points to this message with the text "Recourse Link". Below the message, it says "ALL RESULTS" and "1-8 of 30,500,000 results · [Advanced](#)". The first search result is titled "Welcome to Blair House - Home Page" and includes the URL "blairhouse.org" and a description: "Official site. Built in 1824, and in the service of the nation since 1942, Blair House stands across Pennsylvania Avenue from the White House. It is the President's Guest House, playing host ...". The second result is titled "Blair House - Wikipedia, the free encyclopedia" and includes a description: "Blair House is the official state guest house for the President of the United States. It is located at 1651-1653 Pennsylvania Avenue NW in Washington, D.C., opposite the Old ...". The third result is titled "Welcome to Blair House - Home Page" and includes a description: "Browse the guestbook: Visit specific rooms: From the time Francis Preston Blair and Eliza Blair moved into their new home, it was a politically significant address." and the URL "www.blairhouse.org/home.html".

Recourse
Link

Stemming

The screenshot shows a Bing search results page for the query "vietnamese restaurant maiden lane sf". The page includes a search bar with the query, a "Web" filter, and a list of search results. The results are sorted by relevance, with the top result being "Union Square San Francisco Maiden Lane Cafes and Restaurants". The page also features a sidebar with "RELATED SEARCHES" and "SEARCH HISTORY".

bing
Web

vietnamese restaurant maiden lane sf

ALL RESULTS 1-10 of 37,700 results · Advanced

Union Square San Francisco Maiden Lane Cafes and Restaurants
Union Square Association provides detailed listings of businesses for this unique shopping district
www.unionsquaresf.net/maidenlane.htm · Cached page

Mocca On Maiden Lane San Francisco, CA Restaurant Reviews and Ratings
Mocca On Maiden Lane San Francisco, CA Restaurant Reviews and Ratings ... 107) >
Aux Delices Vietnamese Rstrnt 108) Axis Cafe & Gallery 109) Axum Cafe
www.pubcrawler.com/Template/dsp_restaurant_zoom.cfm/flat/ID=250837 · Cached page

Maiden Lane, San Francisco, CA - Yahoo! Travel
Restaurants; Trip Plans; Photos/Videos; Map; Deals ... Also of interest is the only building in San Francisco designed by Frank Lloyd Wright at 140 Maiden Lane.
travel.yahoo.com/p-travelguide-2733507-maiden_lane_san_francisco-i · Cached page

Vietnamese Restaurants | Metromix New York
9 Maiden Lane - New York (Financial District) ... Restaurants / Vietnamese ... San Francisco; Cincinnati; Houston; Oklahoma City; Seattle
newyork.metromix.com/facets/restaurants/topic.restaurants.cuisine.asian.vietnamese · Cached page

RELATED SEARCHES
Maiden Lane San Francisco
Maiden Lane Advertising
Maiden Lane III LLC
Maiden Lane AIG
Maiden Lane II
Maiden Lane NY
Map for Maiden Lane NYC
77 Maiden Lane

SEARCH HISTORY
vietnamese restaurant san francisco
software development jobs
software development metrics
Similar searches

Equivalent Forms

Blended Results

Abbreviations

The image shows a Bing search results page for the query "un jobs". The search bar at the top contains "un jobs" and a magnifying glass icon. Below the search bar, there are tabs for "Web" and "News". The search results are displayed in a grid format. On the left side, there are sections for "RELATED SEARCHES" and "SEARCH HISTORY". The main content area shows "ALL RESULTS" for "un jobs" with a count of "1-10 of 275,000,000 results" and a link to "Advanced" search options. A "News: un jobs" section is highlighted, featuring a news article snippet with a photo of a smartphone. A callout box labeled "Federated Results" points to this news section. Below the news section, there are social media sharing options (Facebook, Twitter, Messenger, Email) and a link to "Job Vacancies in United Nations and International Organizations". A callout box labeled "Term Expansion" points to this link. Below this link, there is a list of related terms: "Duty Stations", "Internships", "Vacancy", "Latest Tags", "Organizations", "Latest Authors", "Closing Soon", and "Carcasí, Colombia". At the bottom, there is a link to "United Nations System Employment Opportunities" and a snippet of text about "UN Vacancy Announcements: World Health Organization (WHO) Employment site: UNDP Vacancy Announcements: UNDP Ukraine Vacancy Announcements: UN".

bing

un jobs

Web News

RELATED SEARCHES

- Un Job Opportunities
- Un Job Sudan
- Un Job List
- UNjobs Kenya
- UNjobs Africa
- UNjobs Pakistan
- Un Vacancies Sudan
- UNICEF Jobs

SEARCH HISTORY


- iota
- what is an iota
- nap the boom
- nap the boom online
- +name +of +the greek
- cross

See all





Clear all · Turn off

ALL RESULTS 1-10 of 275,000,000 results · [Advanced](#)

News: [un jobs](#)

 Letters to the editor: Immigration, **jobs** creation, unemployment
Contrary to the opinions of letter writer Steve Johnson (July 14), polls show the American public... [San Diego...](#) · 10 hours ago
[CT adds 500 jobs in June WTNH.com](#)
[Competing for jobs Montgomery...](#)
[See today's top stories · See recent tweets · Create alert](#)

Portfolio.com

SHARE  Facebook  Twitter  Messenger  Email

[Job Vacancies in United Nations and International Organizations](#)
Vacancy Announcements from the **United Nations** and other International Organizations, compiled by UNjobs Association of Geneva
[unjobs.org](#) · [Cached page](#)

- Duty Stations
- Internships
- Vacancy
- Latest Tags
- Organizations
- Latest Authors
- Closing Soon
- Carcasí, Colombia

[Show more results from unjobs.org](#)

[United Nations System Employment Opportunities](#)
UN Vacancy Announcements: World Health Organization (WHO) Employment site:
UNDP Vacancy Announcements: UNDP Ukraine Vacancy Announcements: UN

Federated Results

Term Expansion

Term Relaxation

The screenshot shows a Bing search result for the query "what is a diploid chromosome". The search bar contains the query and a magnifying glass icon. Below the search bar, there are tabs for "Web" and "News". The main content area is divided into "RELATED SEARCHES" and "ALL RESULTS".

RELATED SEARCHES:

- 46 Chromosomes
- Diploid Cell
- Diploid Organisms
- Female Chromosomes
- Chromosome Diagram
- Chromosome Number
- Organisms
- Chromosome Number of Animals
- Chromosome Number by Species

SEARCH HISTORY:

- what is a diploid
- where can i get an iphone 4
- +where can i get an iphone 4
- where can i get an iphone
- or and and gates

ALL RESULTS: 1-10 of 306,000 results · [Advanced](#)

Ploidy - Wikipedia, the free encyclopedia
Haploid and monoploid · **Diploid** · Haploidisation · Polyploidy
Ploidy refers to the number of sets of **chromosomes** in a biological cell. Human germ cells (sperm and egg) have one complete set of **chromosomes** from the male or female parent.
en.wikipedia.org/wiki/Diploid · [Cached page](#)

WikiAnswers - What is a diploid chromosome
I'm assuming you mean "What is the **diploid** chromosome number". Basically you have haploid and **diploid**... so in humans 46 is the **diploid** number and 23 is the haploid.
wiki.answers.com/Q/What_is_a_diploid_chromosome · [Cached page](#)

Chromosome - Wikipedia, the free encyclopedia
History · **Chromosomes** in eukaryotes · **Chromosomes** in ...
These tables give the total number of **chromosomes** (including sex **chromosomes**) in a cell nucleus. For example, human cells are **diploid** and have 22 different types of autosome, each ...
en.wikipedia.org/wiki/Chromosome · [Cached page](#)

Chromosomes
Diploid numbers of some commonly studied organisms (as well as a few extreme examples) ... The molecule of DNA in a single human **chromosome** ranges in size from 50 x 10⁶ nucleotide ...
users.rcn.com/jkimball.ma.ultranet/BiologyPages/C/Chromosomes.html · [Cached page](#)

Key term extraction

The screenshot shows a Bing search result for the query "where can i get an iphone 4". The search bar contains the query and a magnifying glass icon. Below the search bar, there are tabs for "Web" and "News". The main content area is divided into "RELATED SEARCHES" and "ALL RESULTS".

RELATED SEARCHES:

- Can Get MySpace School
- Can Get Pregnant My Period
- Can Get Free Music Download
- Can Get Music Codes
- Can Get Rich
- You Can Get Healthy
- Who Can Get Medicaid
- Days You Can Get Pregnant

SEARCH HISTORY:

- +where can i get an iphone 4
- where can i get an iphone 4
- where can i get an iphone
- or and and gates

ALL RESULTS: 1-10 of 685,000 results · [Advanced](#)

News: where can i get an iphone 4

iPhone 4 Overview
... the **iPhone 4** relies on 3rd party Camera applications to **get** the job done. We would have liked a... [infoSync](#) · 4 hours ago
Apple iPhone 4 FAQ: What Happened, What's Next [PC World](#)
5 Reasons To Upgrade To iPhone 4G Touch Screen Phone [Transworld News](#)
See today's top stories · See recent tweets · Create alert

SHARE [Facebook](#) [Twitter](#) [Messenger](#) [Email](#)

How To: Get an iPhone 3G for \$199... Without Contract! (Updated ...)
Step 4: Cancel service. (Optional). For this, you will have to wait exactly 30 days ... How can I get an **iPhone 3G** at the \$199 or \$299 price for an 8gb or 16gb?
www.phoneneews.com/how-to-get-an-iphone-3g-for-199-without-contract-3704

iPhone 4 FAQ: What exactly did Apple do?
Three, if you're still not happy, you can bring your **iPhone 4** back, get a full refund, and apparently get released from your AT&T contract. Four, and not directly related to the ...
www.sfgate.com/.../article.cgi?f=/g/a/2010/07/17/urmidgns852573C40069388002577630055B4DD.DTL · [Cached page](#)

An iPhone Lover's Take On The iPhone 4
By MG Siegler · 531 posts
The processor can achieve 1GHz but it's actually slower than the iPad which suggests

Incorrect intent



Term Relaxation

Incorrect Terminology

bing
Web

Search query: bowker's test of change tutorial

ALL RESULTS 1-10 of 8,480 results · [Advanced](#)

RELATED SEARCHES

- Word Tutorials Tests
- Free Typing Test Tutorial
- Free Aptitude Test Tutorials
- Free Microsoft Excel Tutorial Test
- Advanced Microsoft Excel Tutorial Test
- MVA Learners Permit
- MVA Drivers Test
- Maryland Test

SEARCH HISTORY

- Bowker's test
- Bowker's test tutorial
- Bowker's test of change tutorial
- what is a diploid chromosome
- what is a diploid

ALL RESULTS

- [OQA-43\] Link to SeleniumRC Python tutorial is broken: http ...](#)
Change History; Activity Stream [... Caroline Bowker added a comment - 16/Dec/08 02:56 PM ... then you'll also find the sample Python unit test Since this tutorial is ... [jira.openqa.org/browse/OQA-43](#) · [Cached page](#)
- [Tutorial for Flowol 3 software](#)
... Programming: Anthony Bowker (K.I.T.E.) Tutorial: Rod Bowker (K... Outputs (1 to 6) - 6V bulbs Testing the ... Keep I.T. Easy 2005 Flowol Flowol Tutorial Secondary Activity 7: Changing the speed ... [www.flowol.com/documents/F3Tutorial.pdf](#) · [Cached page](#) · PDF file
- [Brent Bowker - Canada | LinkedIn](#)
Seeking employment in the energy sector in the... · Higher Education ... and encouraged to explore new ideas and test my thinking. Brent Bowker's ... the conservation and demand electricity changes. ... Prepared for and facilitated weekly tutorial ... [www.linkedin.com/pub/brent-bowker/12/467/581](#) · [Cached page](#)
- [Gale - HTML File Page](#)
Guided Tutorials; MARC Records; Navigation Guides; Product Reviews ... duplication is partly due to sloppy work, partly to Bowker's ... The test search for verifying the completeness of the ... [www.gale.cengage.com/.../archive/200007/bowker.html](#) · [Cached page](#)

bing
Web

Search query: Bowker's test

ALL RESULTS 1-10 of 161,000 results · [Advanced](#)

RELATED SEARCHES

- P S a Test
- U S Test
- Cayman S Test
- S Endorsement Test
- Test Speed
- S Online
- Fun Online Tests
- SMAP

SEARCH HISTORY

- Bowker's test tutorial
- Bowker's test of change tutorial
- what is a diploid chromosome
- what is a diploid
- where can i get an iphone 4

We're including results for **bowker** test. Do you want results only for **Bowker's** test ?

ALL RESULTS

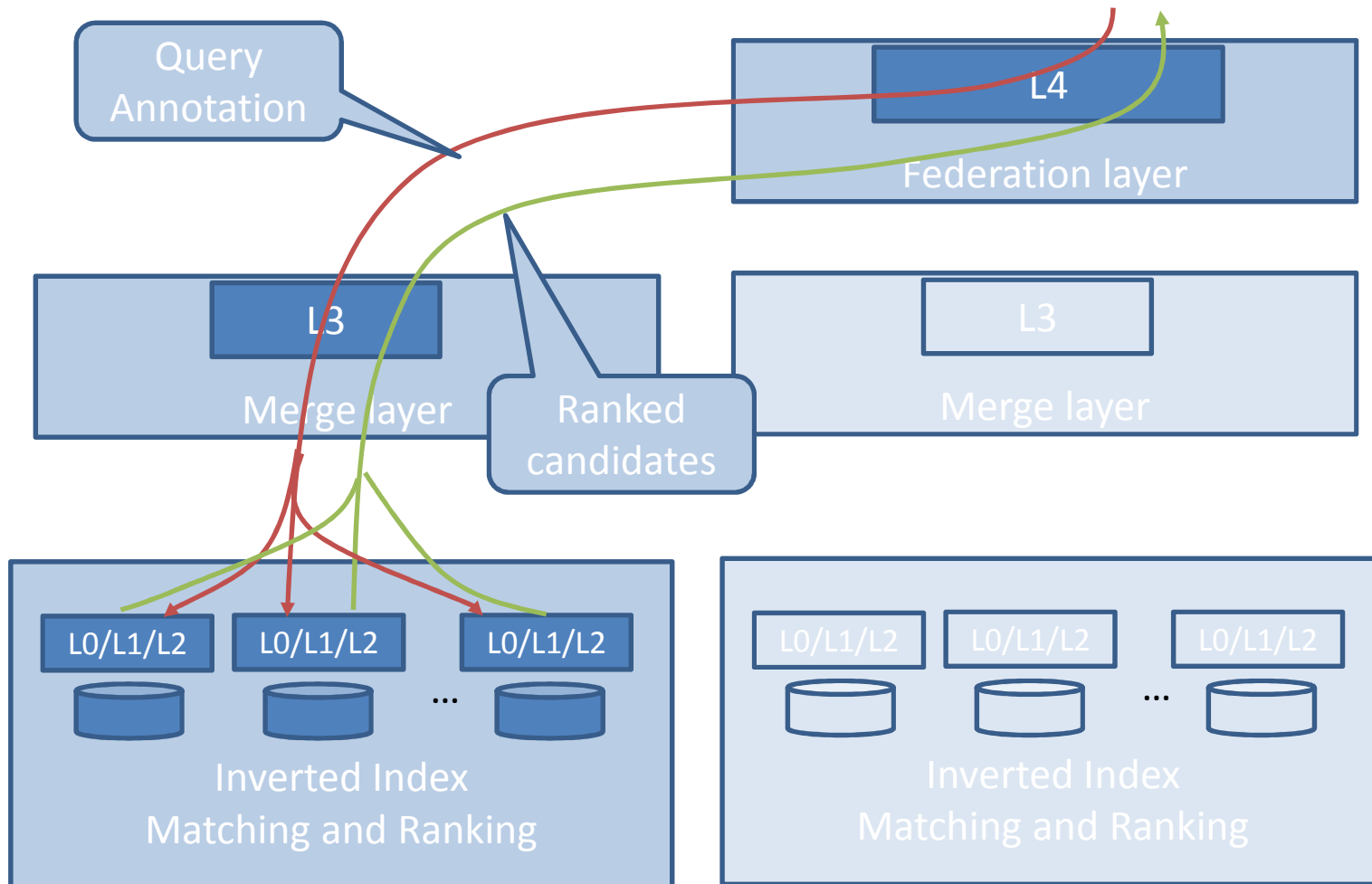
- [\(Page 10 of 11\) - A Weighted Test of Marginal Homogeneity authored by ...](#)
toward calculus that occurred after students completed a unit on business calculus. The authors used Bowker's test to assess this change in attitude. In the present study, the ... [www.allacademic.com/pages/p275428-10.php](#) · [Cached page](#)
- [A Weighted Test of Marginal Homogeneity](#)
Bowker's test is a hypothesis test used to assess change in matched paired data with multinomial responses. A criticism of Bowker's test is that it does not account for the ... [www.allacademic.com/metal/p275428_index.html](#) · [Cached page](#)
- [Maxwell's test, McNemar's test, Kappa test](#)
It is then called the McNemar-Bowker test of symmetry. It tests for symmetry around the diagonal of the table. The general McNemar statistic tests for asymmetry in the distribution ... [rimarcik.com/en/navigator/z-nominal.html](#)
- [Bowker](#)
Teacher: Bowker ... dent To Test [www.thatquiz.org/tq/classpage?005235689ae34aa](#) · [Cached page](#)
- [Newscow - Search Results - Bowker](#)

Relaxed Query

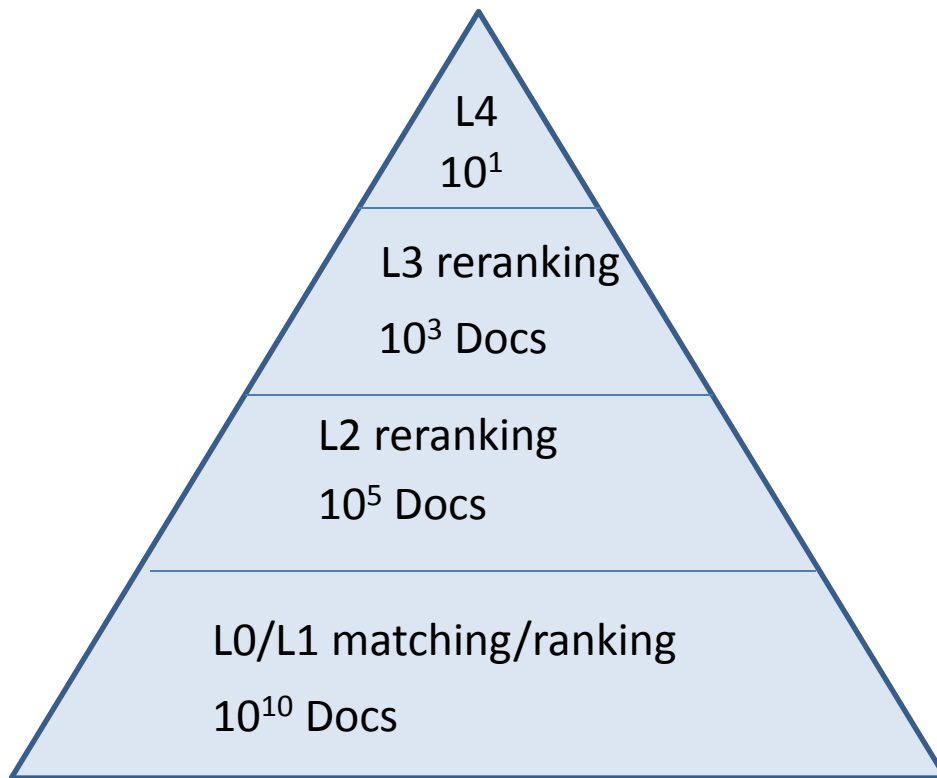
Key Concepts

- Win/Loss ratios
 - Automated reformulations are errorful
 - Wins are queries whose results improve
 - Losses are queries whose results degrade
 - Related to precision
 - But not all valid reformulations change results
- Pre vs Post result analysis
 - Query alternatives generated pre-results
 - Blending decisions are post results

Query Evaluation



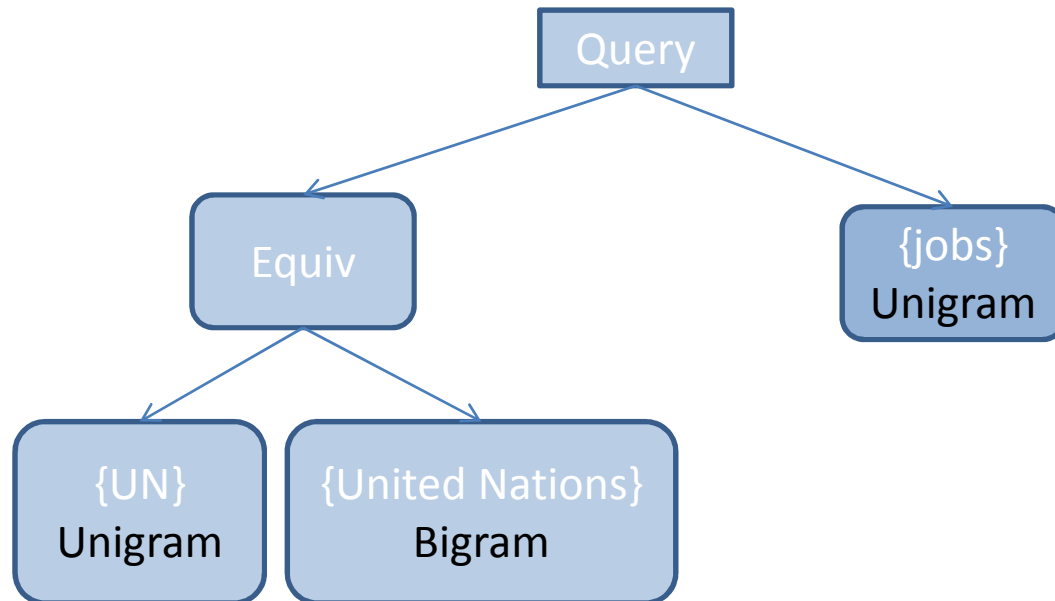
Matching and Ranking



- Sequence of select, rank, prune steps
- L0: Boolean logic
- L1: IR score
- L2/L3/L4: Machine learned

Learning to Rank Using Gradient Descent,
Burges et al., ICML 2005

Query Annotation



- NLP query annotation
 - Offline analysis
- Ambiguity preserving
 - Multiple interpretations
- Backend Independent
 - Shared
- Structure and Attributes
 - Syntax and semantics

Query Planning

[{un | “united nations”} jobs] →
L3-merge(L2-rank([un jobs]),
L2-rank([“united nations” jobs]))

or

[{un | “united nations”} jobs] →
L3-cascade(threshold, L2-rank([un jobs]),
L2-rank([“united nations” jobs]))

Design Considerations

- Efficiency
 - Query plan can generate multiple backend queries
 - Merged in L3
 - Some queries are cheaper than others
 - Query reduction can improve performance
- Relevance
 - L3 merging has maximal information
 - But is costly
- Multiple query plan strategies
 - Depending on query analysis confidence

Query Analysis Models

- Noisy Channel Model

$$\arg \max \{P(\textit{rewrite} \mid \textit{query})\} = \\ \arg \max \{P(\textit{rewrite})P(\textit{query} \mid \textit{rewrite})\}$$

Language
Model

Translation
Model

- Example: Spelling
 - Language Model: Likelihood of the correction
 - Translation Model: Likelihood of the error occurring

Language Models

- Based on Large-Scale Text Mining
 - Unigrams and N-grams
 - Probability of query term sequence
 - Favor queries seen before
 - Avoid nonsensical combinations
- 1T n-gram resource
 - **Multi-Style Language Model for Web Scale Information Retrieval**
Wang et al., SIGIR 2010
 - **SCOPE: Easy and Efficient Parallel Processing of Massive Data Sets**
Chaiken et al., PVLDB 2008

Translation Models

- Training set of aligned pairs
 - E.g. misspelling/correction; surface/stemmed
- Query log analysis
 - Session reformulation → associated queries
 - Co-click → associated queries
 - Manual Annotation
- Many application-specific models
 - **Random Walks on the Click Graph**
Craswell et al., SIGIR 2007
 - **Context sensitive synonym discovery for web search queries**
Wei et al., CIKM 2009
 - **Context sensitive stemming for web search**
Peng et al, SIGIR 2007

Summary

- Query Understanding is critical to Web Search
 - Affects most queries
 - Can radically improve results
- Trade-off between relevance and efficiency
 - Rewrites can be costly
 - Win/loss ratio is the key metric
- Especially important for tail queries
 - No meta-data to guide matching and ranking