

Multilingual Information Access: The User's Perspective

Douglas W. Oard

College of Information Studies and
Institute for Advanced Computer Studies
University of Maryland, USA

Outline

- Deconstructing CLEF
- The user's perspective
- User interface design
 - Transparent query translation
 - Selection and examination
- A modest proposal

Deconstructing CLEF

- What user process is modeled by the CLEF evaluation paradigm?
- Answer: **Interactive** retrospective search
 - User presents a previously unseen query
 - Expectation over “random” queries
 - System produces a ranked list
 - User scans the list from the top until satisfied
 - Expectation over potential stopping points

The User's Perspective: IR is Only Part of the Story

- “Pull” Information systems
 - Search
 - Browse
- “Push” information systems
 - Filtering
 - Advertising
- Other people
 - Gatekeepers

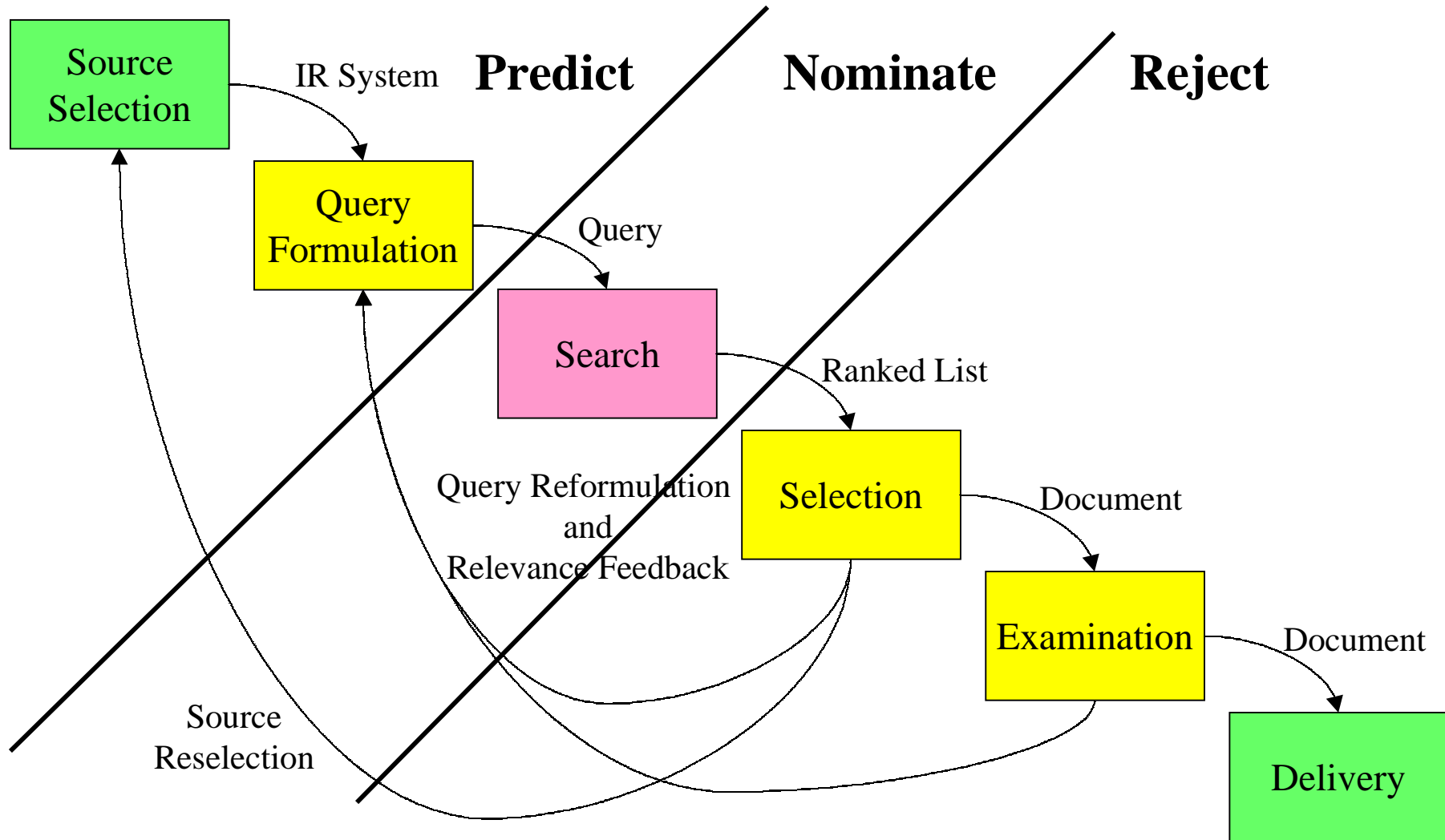
Information Access Tasks

- Real-time alerting
 - Example metric: Error rate
- Precision-biased retrospective search
 - Example metric: Mean precision at 10 documents
- Recall-biased retrospective search
 - Example metric: Mean R-precision
- Known-item retrieval
 - Example metric: Mean inverse rank
- Exploration
 - Example metric: Aspectual recall

Some Fundamental Principles of Interactive IR System Design

- Machines are fast, but stupid
 - Partition the task so that operations requiring speed are assigned to the machine
- Humans are smart, but slow
 - Design the user interface to optimize the users performance at tasks that are beyond the capability of the machine

Supporting the Search Process



Transparent Query Translation

Enter Query:
bridge and tunnel construction for the Beijing-Kowloon railroad

Select term Meaning:

bridge	桥梁	a structure carrying a pathway or roadway over
bridge	跨越	to make a bridge over or across□
tunnel	地下道	a subterranean gallery (as in a mine)□
tunnel	隧道	a covered horizontal passageway through or unde
construction	构造	the act or result of construing, interpreting,
construction	建造	the process, art, or manner of constructing som
construction	建筑物工程	the construction industry□

No Translation for:
Kowloon

search on selected terms

Searching on these terms:
Beijing railroad bridge(a structur . . .) tunnel(a covered . . .) construction(the proces . . .)

Search Results: (Select document to retrieve)

1. ANDES Partners Corporation
2. Building Material Directory
3. Chinese Business World
4. Economics of Networks Internet Site
5. Hong Kong Construction Industry

Enter Query:

Select term Meaning:

bridge	ÇÄÄ°	a structure carrying a pathway or roadway over
bridge	¿çÓ½	to make a bridge over or across <input type="checkbox"/>
tunnel	μθĪÀμÀ	a subterranean gallery (as in a mine) <input type="checkbox"/>
tunnel	ĒipÀ	a covered horizontal passageway through or unde
construction	¹¹Óì	the act or result of construing, interpreting,
construction	½ĒÈ	the process, art, or manner of constructing som
construction	½"ÖpĪi¹¼¹Ī	the construction industry <input type="checkbox"/>

No Translation for:

Kowloon

Searching on these terms:

Beijing railroad bridge(to make a . . .) tunnel(a covered . . .) construction(the proces . . .)

Search Results: (Select document to retrieve)

1. ANDES Partners Corporation
2. Building Material Directory
3. Chinese Business World
4. Economics of Networks Internet Site
5. Hong Kong Construction Industry

Selection and Examination

Query bridge and tunnel construction for the Beijing-Kowloon Railroad

Title coal eastward

Date 99/07/07

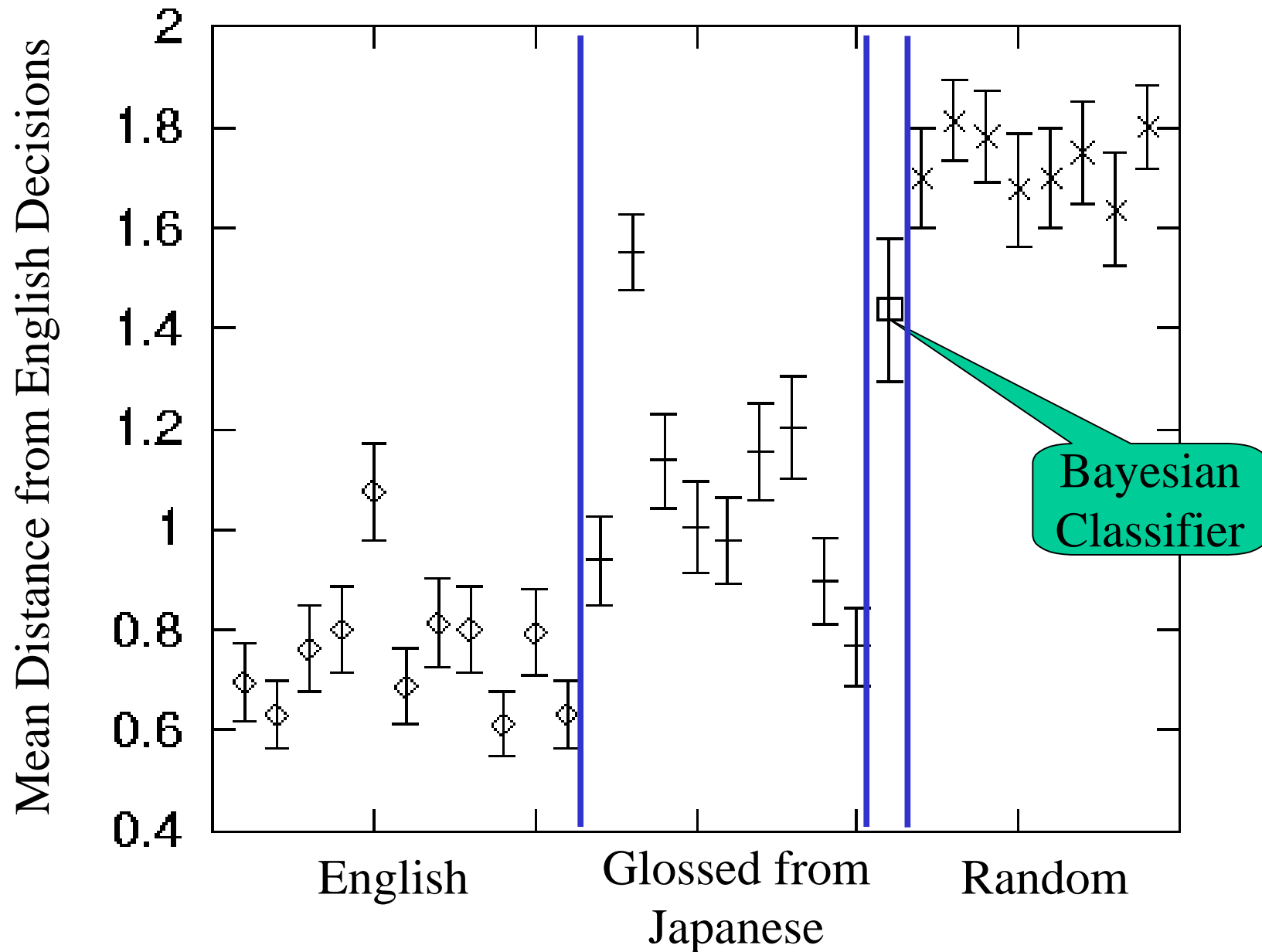
西煤东运大通道神华工程生命线 朔黄铁路优质高效加快建设 神肃段明年10月1日前率先开通运煤 990707 西煤东运大通道神华工程生命线 朔黄铁路优质高效加快建设 神肃段明年10月1日前率先开通运煤 本报讯记者严冰报道：我国西煤东运的第二大通道、举世瞩目的跨世纪工程神华工程的生命线——朔黄铁路正在广大建设者手中，优质、高效地建设成为具有我国九十年代末科技水平的现代化铁路。朔黄铁路西起神朔线的神池南站，东至黄骅港的港前站，全长588公里，总投资193亿元，由神华集团、铁道部、河北省、山西省共同出资建设，除资本金50.3亿元外，主要资金来源于国家开发银行和日本海外协力基金贷款。目前已开工建设的神（神池南）肃（肃宁北）段正线全长419.6公里，为双线电气化铁路，国家一级干线，概算投资146.2亿元。线路横穿恒山、云中山脉和忻定盆地，沿滹沱河谷穿越太行山进入华北平原。沿线地质情况极为复杂，西段山高谷深，有在我国双线长大隧道中居第2位的长梁山隧道和位居第4和第5位的寺铺尖隧道、水泉湾隧道，线路15次跨越滹沱河；有墩高达61.5米的庄里、红山崖、滴流磴3座悬灌梁大桥，主跨最大达80米。东段河网交错，公路纵横，人口稠密，跨越运输繁忙的京广、京九铁路干线和京深高速公路、107国道等公路干线，1公里以上的特大桥就有5座，跨京深高速公路特大桥（667米）主跨以64米下承钢桁梁跨越公路，为我国铁路之仅有。朔黄铁路的工期目标是2000年5月31日神肃段铺通，10月1日与京九接通运煤，2002年与黄骅港同期配套建成。目前，神肃段已完成路基土石方5000多万立方米，占设计总量的90%多，桥梁工程完成3万多双成桥米，占设计总量的90%多，隧道完成5万多成洞米，占设计总量的70%多，设计的77座隧道已贯通60座；正线铺轨完成200多公里，占设计总长的30%多，站线铺轨完成40多公里，占设计总长的20%多，架梁完成500多孔，占设计总量的20%多。朔黄铁路已开工的各项主体工程均由中标单位承担，不准分包，从制度上保证了工期，质量始终处于可控状态。由于施工单位素质高，能力强，施工管理严格，在朔黄铁路建设中创造了不少奇迹。全线工程质量良好，已完成的工程质量合格率100%，优良率93%。据了解，朔黄铁路建设者提出的质量目标是，路基工程要赶超重载铁路大秦线，桥、隧等结构工程要赶超京九、南昆线，站后工程要面向21世纪。《海外版》（1999年07月07日第2版）人民日报社版权所有，未经授权禁止复制或建立镜像。

info@peopledaily.com.cn 电话：(010)65092993 (010)65091079

User Study Results

- **Important:** Strong support for interactive relevance judgments can make up for less accurate nominations
 - [Hersh et al., 2000]
- **Practical:** Interactive relevance judgments based on imperfect translations can beat fully automatic nominations alone
 - [Oard and Resnik, 1997]

Manual Cross-Language Classification



Is Off-the-Shelf MT Enough?

- Speed: An order of magnitude too slow
 - Goal: translate a screen in 100 ms
- Accuracy: Disfluencies cause slow recognition
 - Possible solution: display alternative word choices
- Cost: Developing language pairs is expensive
 - Compromise: Balance accuracy and cost

State-of-the-Art Machine Translation

shipment 大通道神 China engineering life line new moon yellow **railroad** high grade highly effectively will speed up **construction** god 肃 section next year in October 1 other day takes the lead to clear West coal eastward shipment 大通道神 China engineering life line new moon yellow railroad high grade highly effectively will speed up construction god 肃 section next year in October 1 other day takes the lead to clear transports coal This 报讯 reporter 严冰 report: West our country coal eastward shipment second big channel, whole world focuses attention on steps in century engineering god China engineering life line —— new moon yellow **railroad** broad constructor hand, high grade, highly effectively constructs into has our country the '90s end science and technology horizontal modernization **railroad**. West new moon yellow **railroad** god new moon line Shenchi south station, east to 黄骅 port port advance party, span 588 kilometers, total budgetary estimate invests 193 hundred million Yuan, province, Shanxi Province together invests construction by north god China group, ministry of railways, river, eliminates capital capital 50. Outside 3 hundred million Yuan, main fund originates at country develops bank and Japan overseas unites efforts fund loan. Presently has begun construction god (south Shenchi) 肃 (肃宁 north) section main line span 419. 6 kilometers, are double thread electric railway, country one level of skeleton lines, budgetary estimate invests 14 6. 2 hundred million Yuan. Line crosses 恒山, in the clouds sierra and 忻 decides trough, passes through Mount Taihang along Hutuo River river valley enters North China plain. Geology situation is along the route most complex, west section mountain high valley is deep, has occupies 2nd long Liangshan **tunnel** in our country double thread long big tunnel and is situated 4th and 5th Buddhist temple shop point **tunnel**, 水泉 bay **tunnel**, line 15 time surmounts Hutuo River river; Has pillar reaches as high as 61. In 5 meters villages, red cliff, dribbles 磴 3 hangs fills Liang big bridge, host steps most greatly reaches 80 meters. East section network of waterways interlocks, road is vertical and horizontal, population is dense, surmounts transports loses busy **Beijing** is broad, **Beijing** nine railroads road skeleton lines such as skeleton

Proposal: Interactive CLEF

- Collection
 - Single original document language
 - Users get translations in their preferred language
 - Everyone uses ranked lists from two systems
 - Best of CLEF 2000 and a below-median system
- Metric
 - Aspectual recall in a fixed time
 - Requires additional assessment effort

Conclusion

- Automatic techniques will eventually reach a diminishing return on investment
- An interactive track at CLEF can add insight
 - How sensitive are users to weak automatic search?
 - How should the selection and examination interfaces be designed?

For Further Information

- June 2000 Interactive CLIR workshop
<http://www.clis.umd.edu/conferences/hci100/>
- Interactive CLIR demo
<http://tides.umiacs.umd.edu>
- Cross-language IR resources
<http://www.clis.umd.edu/dlrg/clir/>

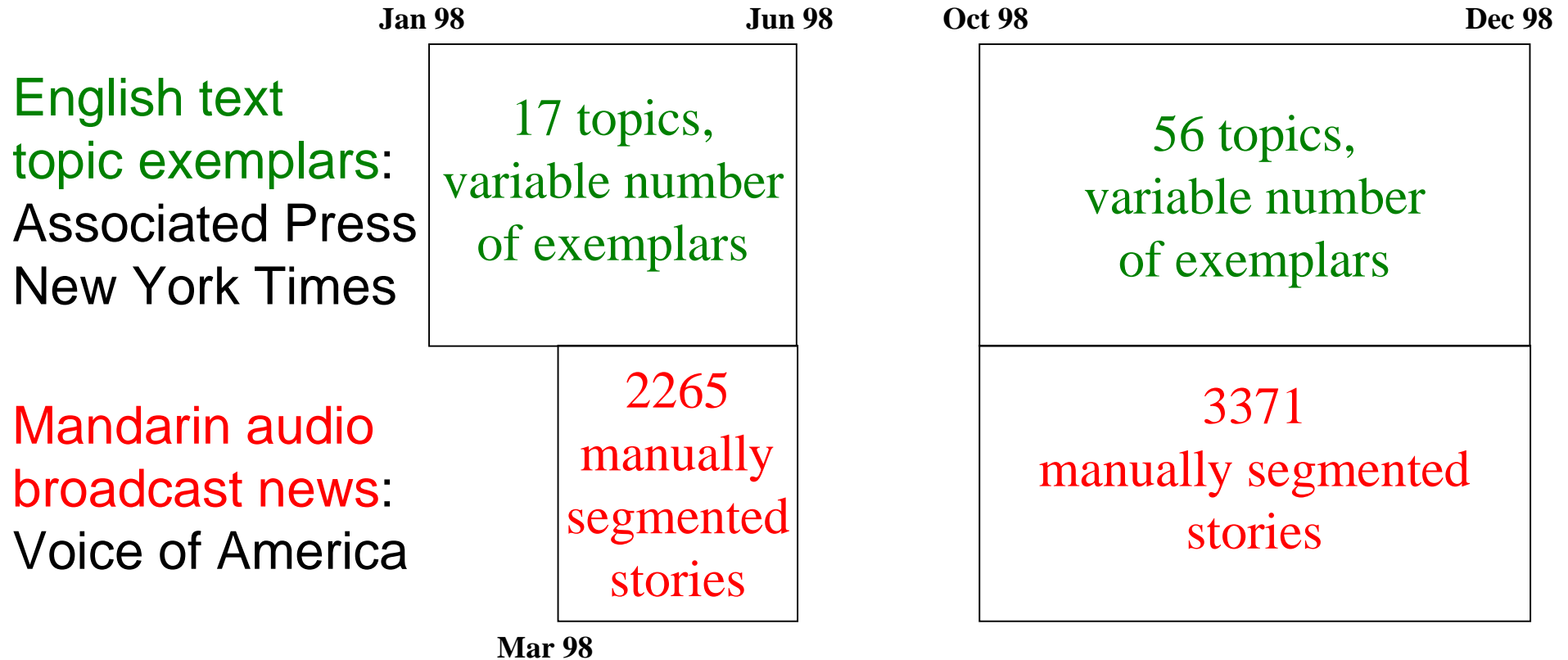
And now for something
completely different...

Cross-language Speech Retrieval

Topic Detection and Tracking: Cross-Language Speech Retrieval

**Development
Collection: TDT-2**

**Evaluation
Collection: TDT-3**



Exhaustive relevance assessment based on event overlap

Hopkins MEI Workshop Results

