

Metasearch

ISU535

Prof. Aslam

Search Engines

- Provide a ranked list of documents.
- May provide relevance scores.
- May have performance information.

Search Engine: Alta Vista

The screenshot shows the Alta Vista search engine interface. At the top, there is a navigation bar with the Alta Vista logo and links for Search Home, Comparison Shop, Channels, Rewards, and Email & Tools. Below this is a search bar containing the text "chili peppers". To the right of the search bar is a "Search" button. Below the search bar, there are several sections: "AltaVista Recommends" with links for music and entertainment, and "Find Results In:" with tabs for Products, News, Business, Web Pages, Images, MP3/Audio, Video, and Directories. The search results are listed below, starting with "chili peppers" and "1. Red Hot Chili Peppers".

File Edit View Go Communicator Help
Bookmarks Location: <http://www.altavista.com/cgi-bin/query?q=chili+peppers&kl=XX&pg=q&Translate=on&search.x=18&search.y=11> What's Related

alta vista: Search Home Comparison Shop Channels Rewards Email & Tools
Sign Up! AltaVista Members Sign In Member Center My AltaVista

alta vista: SHOPPING Gift of Learning Sweepstakes Click Here!
Click here to find out more!

Search for: chili peppers any language Search Help | Customize Settings | Family Filter is off

Search within these results Example: +skin +cancer -ultraviolet

Related Searches:
- red hot chili peppers - red hot chili peppers
- red hot chili peppers - red hot chili peppers
- chili peppers tabs - chili peppers links

AltaVista Recommends

* Music with: [chili peppers](#)
* AltaVista Entertainment: [Music](#)

Find Results In: 58,832 pages found.
Products News Business Web Pages Images MP3/Audio Video Directories

chili peppers - Click here for a list of Internet Keywords related to **chili peppers**

See editor-selected sites in: [The Red Hot Chili Peppers](#) | [Red Hot Chili Peppers Band Members](#) | [Red Hot Chili Peppers Fan Dedications](#) | [Red Hot Chili Peppers Blog and Discographies](#) | [Red Hot Chili Peppers News and Zines](#) | [Red Hot Chili Peppers Multimedia](#)

1. [Red Hot Chili Peppers](#)
This web site was created for FREE at [www.homestead.com](#). Visit [www.homestead.com](#) to get your free web site - no programming required.
URL: [www.homestead.com/musiccorner/hotchilipeppers.html](#)
[Translate](#) [More pages from this site](#) [Related pages](#)

2. [Abbey Road: Red Hot Chili Peppers Fansite](#)
For fans of pre-1992 Red Hot Chili Peppers Funk Style!..
URL: [members.xoom.com/old_rhcp/index.html](#)
[Translate](#) [Related pages](#) Facts about: [Xoom.Com Inc](#)

3. [dotmusic - Single Review - Red Hot Chili Peppers - "Otherside" \(WEA\)](#)
Typically groovy bassist Flea fuels the momentum
URL: [www2.dotmusic.com/reviews/Singles/Januar...eviews12638.asp](#)
[Translate](#) [Related pages](#)

4. [Red Hot Chili Peppers](#)
Buy Californication Now. Enter your email address to receive information about The Red Hot Chili Peppers. Email: Go to the RED HOT CHILI PEPPERS...
URL: [www.wbr.com/chilipeppers/index.html](#)
[Translate](#) [Related pages](#) Facts about: [Warner Brothers S...](#)

5. [Hot Chili Peppers Homepage](#)

alta vista: SHOPPING Computer Hardware Gifts & Flowers Furniture
alta vista: CAREERS Post Resume Career Advice Search Tips

Search Engine: Ultraseek

The screenshot displays the Ultraseek search engine interface within a web browser window. The browser's address bar shows the URL: `http://ultraseek.dartmouth.edu:8765/query.html?qt=chili+peppers&rq=0&vs=0&st=1&nh=10&lk=1&rf=0&rq=0`. The page header features the Dartmouth logo (est. 1769) and navigation links for HOME, INDEX, HELP, and SEARCH. The search bar contains the query "chili peppers" and includes a "GO" button. Below the search bar, there are options to "Click to expand search:" with checkboxes for "Select all", "chile", and "chilli".

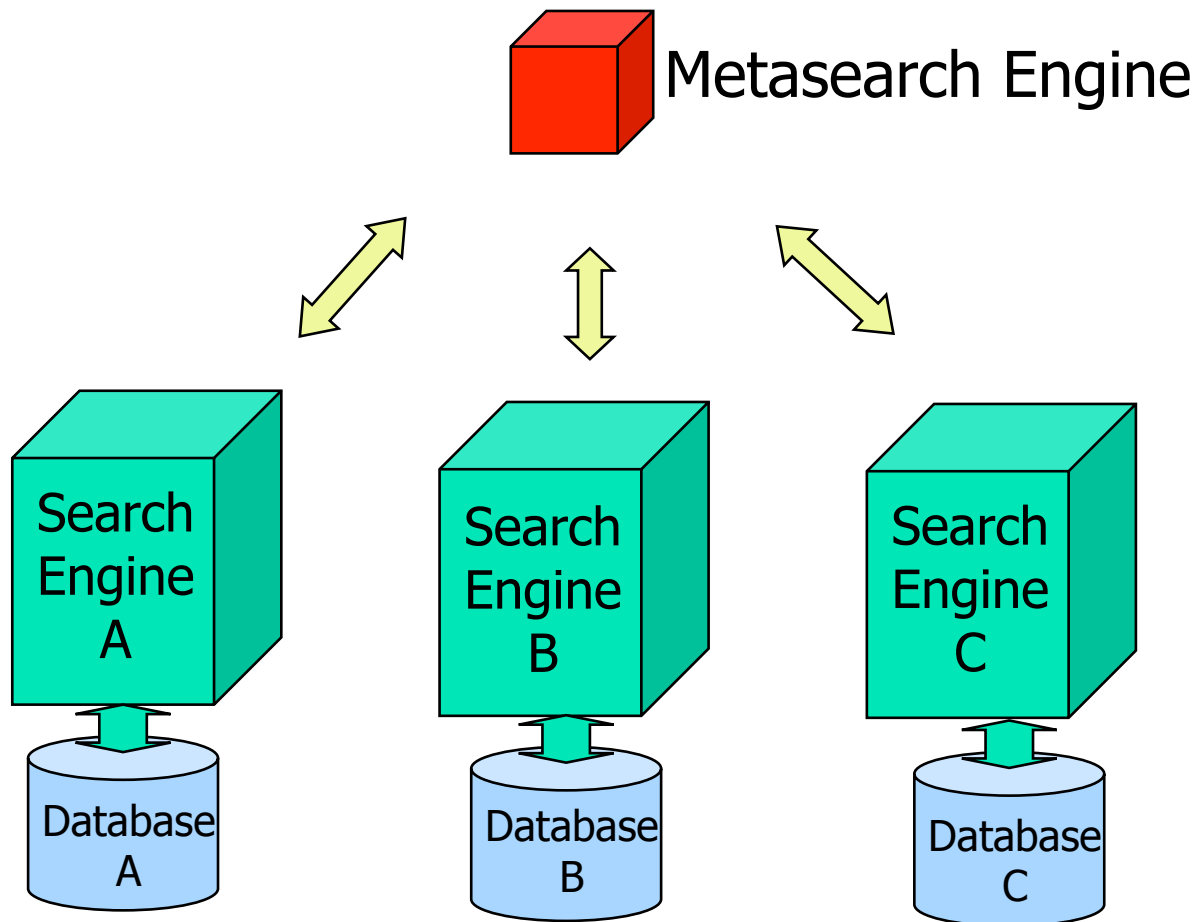
The search results section shows "144 results found, sorted by relevance" and "Document count: chill (46) peppers (129) chili peppers (6)". The results are displayed as a list of items, each with a title, a brief description, a date, and a similarity percentage. The first result is "Home Plate menu 11/5 to 11/10" with a 74% similarity. The second result is "my babu" with a 68% similarity. The third result is "Home Plate Menu 11/12-17" with a 65% similarity.

Other features include "Other Searches" with links to Dartmouth Name Directory (DND), Dartmouth Library Catalog, and Dartmouth College Information System (DCIS). A "Help Using the Search Engine" sidebar provides links for "Searching smart:", "Getting listed:", and "FAQ:". The browser's status bar at the bottom shows "100%" zoom and various system icons.

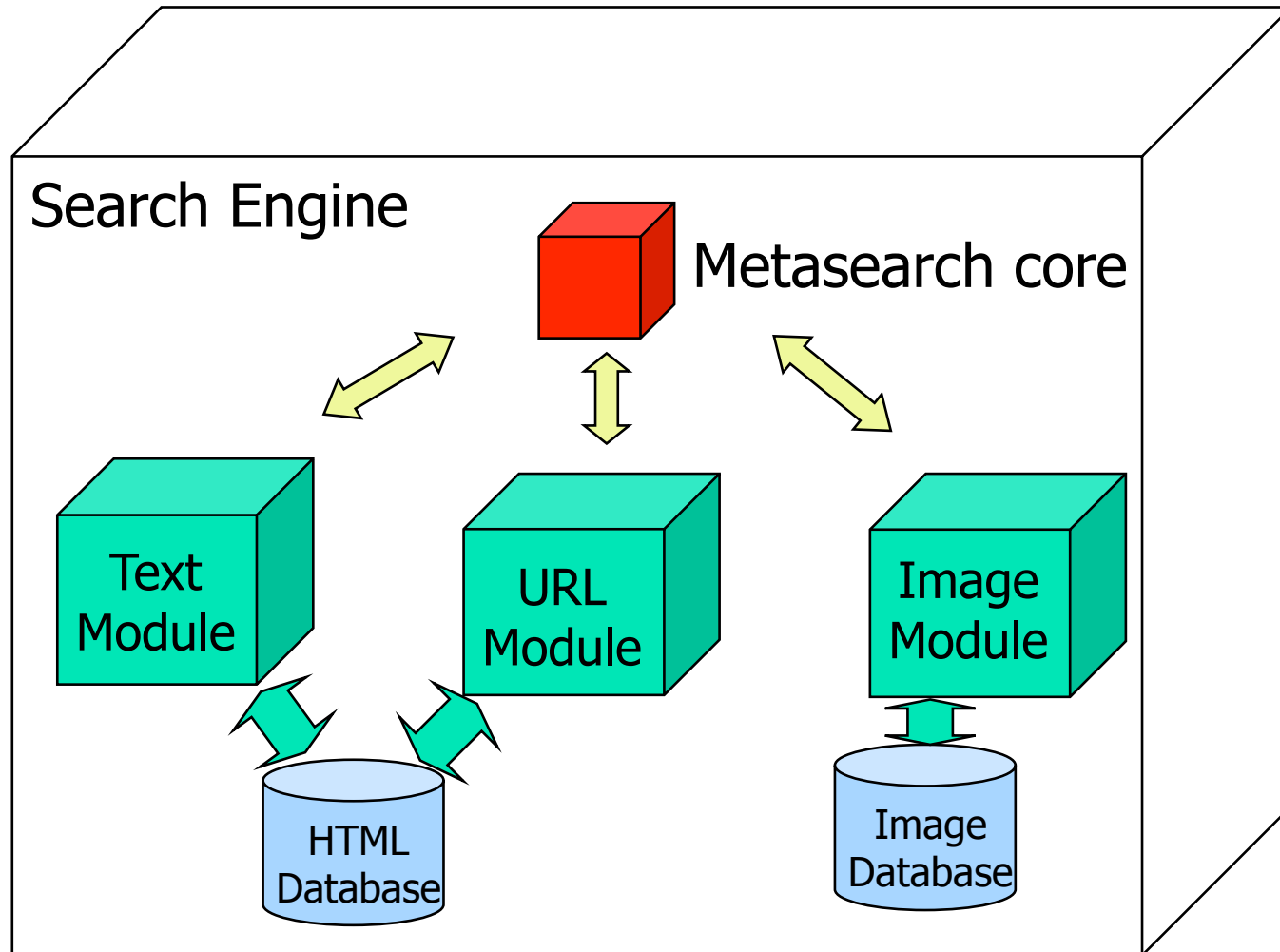
Search Engine: inq102 TREC3

```
Queryid (Num):          50
Total number of documents over all queries
  Retrieved:            50000
  Relevant:               9805
  Rel_ret:              7305
Interpolated Recall - Precision Averages:
  at 0.00               0.8992
  at 0.10               0.7514
  at 0.20               0.6584
  at 0.30               0.5724
  at 0.40               0.4982
  at 0.50               0.4272
  at 0.60               0.3521
  at 0.70               0.2915
  at 0.80               0.2173
  at 0.90               0.1336
  at 1.00               0.0115
Average precision (non-interpolated)
for all rel docs (averaged over queries)
  0.4226
Precision:
  At    5 docs:        0.7440
  At   10 docs:        0.7220
  At   15 docs:        0.6867
  At   20 docs:        0.6740
  At   30 docs:        0.6267
  At  100 docs:        0.4902
  At  200 docs:        0.3848
  At  500 docs:        0.2401
  At 1000 docs:        0.1461
R-Precision (precision after R
(= num_rel for a query) docs retrieved):
  Exact:                0.4524
```

External Metasearch



Internal Metasearch



Metasearch Engines

- Query multiple search engines.
- May or may not combine results.

Metasearch: Dogpile

The screenshot shows a web browser window with the address bar containing the URL: <http://search.dogpile.com/textis/search?q=chili+peppers&geo=no&fs=web&av=custom&engines=goto&engines=Looksmart&engines=tl>. The page features the Dogpile logo and navigation links. A search bar contains the query '+chili +peppers'. Below the search bar, there are several search results from the Looksmart engine, including links to Amazon.com, Electric Library, and various recipe and music sites. The page also includes a 'Dogpile Suggests' section and a 'ClassMates.com' advertisement.

File Edit View Go Communicator Help

Bookmarks Location: <http://search.dogpile.com/textis/search?q=chili+peppers&geo=no&fs=web&av=custom&engines=goto&engines=Looksmart&engines=tl> What's Related

Home
Custom Search
Dogpile Remote
Search at Home
Help with Syntax
Tell a Friend
Go Shopping

 **DOGPILE**
Web Metasearch Results

 Visit the Ask Our Teachers column at SmarterKids.com.

Spend web time wisely. Shop smarter. Hear My Mail from Weblev! Find old classmates here! ClassMates.com

• Buy books about "chili peppers" at Amazon.com • Search for "chili peppers" on Electric Library

Are you looking for:

Red Hot Chili Peppers	Peppers	Chili	Hot Peppers
Red Hot Chili Peppers Tabs	Canning Recipes Peppers	Red Hot Chili Peppers Lyrics	Salsa Recipes Free

 [ClassMates.com](#): Remember your promise to "keep in touch" at graduation? With over 8.5 million registered high school alumni, chances are you'll find your old friends.

Search engine: [Looksmart](#) found 117 results.
The query string sent was [+chili +peppers](#)

- [The Red Hot Chili Peppers](#)
Find photos, lyrics, updates, tour info, and news on alternative-funk-rock band the Red Hot Chili Peppers.
[Looksmart category - Red Hot Chili Pepper](#)
- [Red Hot Chili Peppers Audio and Video](#)
Watch videos and listen to music by this rock/funk band.
[Looksmart category - Red Hot Chili Peppers](#)
- [Chili and Hot Sauces](#)
Shop for mouth-burning chili sauces, Tabasco, hot salsas and other pepper-inspired sauces.
[Looksmart category - Chili & Hot Sauces](#)
- [Chili and Hot Sauces](#)
Find chili and other hot sauce recipes, including salsas, dips, spices, and rubs, and visit the Pepper Fool.
[Looksmart category - Chili & Hot Sauces](#)
- [Red Hot Chili Peppers - Screens and Themes](#)
Promotional screensaver for the funk-rock band features falling chili peppers.
[LookSmart category - Red Hot Chili Peppers Multimedia](#)

[Next set of results from Looksmart](#)

Search engine: [GoTo.com](#) found 10 or more results.

Metasearch: Metacrawler

The screenshot shows a web browser window with the Metacrawler search engine interface. The search bar contains the text "chili peppers" and the search button is labeled "Search". Below the search bar, there are several search results for "chili peppers". The first result is "Red Hot Chili Peppers" with a link to "Red Hot Chili Peppers Tabs". Other results include "Peppers", "Chilli", "Hot Peppers", "Canning Recipes Peppers", "Red Hot Chili Peppers Lyrics", and "Salsa Recipes Free". There is also a section for "Discovery Health" with a "Find!" button and a "Get Auto Insurance" advertisement. The search results are displayed on a yellow background with a navigation bar at the bottom. The browser's address bar shows the URL: <http://search.metacrawler.com/crawler?general=chili+peppers&method=0&redirect=web&pp=20&hp=10®ion=0&timeout=0&sort=>

Metasearch: Profusion

The screenshot displays the Profusion metasearch engine interface. At the top, there is a navigation bar with 'File', 'Edit', 'View', 'Go', 'Communicator', and 'Help' menus. Below this is a search bar with the text 'chili peppers' and a 'Web Search' button. The search results are listed below, showing a ranking of 1-10 of 63 results. The results include links to various websites, summaries, and URLs. The results are as follows:

Rank	Title	Summary	URL
1.0000	Red Hot Chili Peppers	Summary: This web site was created for FREE at www.homestead.com. Visit www.homestead.com to get your free web site - no programming required.	URL: http://www.homestead.com/musiccorner/hotchilipeppers.html (AllVisits)
0.9667	Abbey Road: Red Hot Chili Peppers FanSite	Summary: For fans of pre-1992 Red Hot Chili Peppers Funk Style!	URL: http://members.xoom.com/old_rhcp/index.html (AllVisits)
0.9333	dotmusic - Single Review - Red Hot Chili ...	Summary: Typically groovy bassist Flea fuels the momentum	URL: http://www2.dotmusic.com/reviews/Singles/January2000/reviews12638... (AllVisits)
0.9000	Red Hot Chili Peppers	Summary: Buy Californication Now. Enter your email address to receive information about The Red Hot Chili Peppers. Email: Go to the RED HOT CHILI PEPPERS...	URL: http://www.wbr.com/chilipeppers/index.html (AllVisits)
0.8781	Chili Pepper Rub for Brisket Recipe	Summary: Chili Pepper Rub for Brisket recipe.	URL: http://bbq.about.com/library/rec/br70923a.htm (About)
0.8667	Red Hot Chili Peppers - Screens and Themes	Summary: Promotional screensaver for the funk-rock band features falling chili peppers. More like this: Red Hot Chili Peppers Multimedia	URL: http://www.screensandthemes.com/Previews-Download/Screen-Savers/... (LookSmart)
0.8667	Briguy's Red Hot Chili Peppers & Jeha ...	Summary: Bootlegs of great classic rock, the Red Hot Chili Peppers, and John Fusciante. I've got stuff from both the Californication era and the pre-heroin...	URL: http://members.aol.com/californiandra/boots.html (AllVisits)
0.8342	Chili Pepper smoked Trout	Summary: Chili Pepper smoked Trout recipe.	URL: http://bbq.about.com/library/rec/br70916a.htm (About)
0.8333	Chili Peppers	Summary: Did You Know...? Home The Legend Legend Form A few items for those interested in chili peppers: Scoville Test For Capsaicin - A Thermal...	URL: http://www.cberseacc.com/knowledge.htm (AllVisits)
0.8000	Hot Chili Peppers Homepage	Summary: Hot Chili Peppers Homepage (German)...	

Outline

- ✓ Introduce problem
- Characterize problem
- Survey techniques
- Upper bounds for metasearch

Characterizing Metasearch

- Three axes:
 - common *vs.* disjoint database,
 - relevance scores *vs.* ranks,
 - training data *vs.* no training data.

Axis 1: DB Overlap

- High overlap
 - data fusion.
- Low overlap
 - collection fusion (distributed retrieval).
- *Very different techniques for each...*
- Today: data fusion.

Classes of Metasearch Problems

	no training data	training data
ranks only	Borda, Condorcet, rCombMNZ	Bayes
relevance scores	CombMNZ	LC model

Outline

- ✓ Introduce problem
- ✓ Characterize problem
- Survey techniques
- Upper bounds for metasearch

Classes of Metasearch Problems

	no training data	training data
ranks only	Borda, Condorcet, rCombMNZ	Bayes
relevance scores	CombMNZ	LC model

CombSUM [Fox, Shaw, Lee, et al.]

- Normalize scores: $[0,1]$.
- For each doc:
 - sum relevance scores given to it by each system (use 0 if unretrieved).
- Rank documents by score.
- Variants: MIN, MAX, MED, ANZ, MNZ

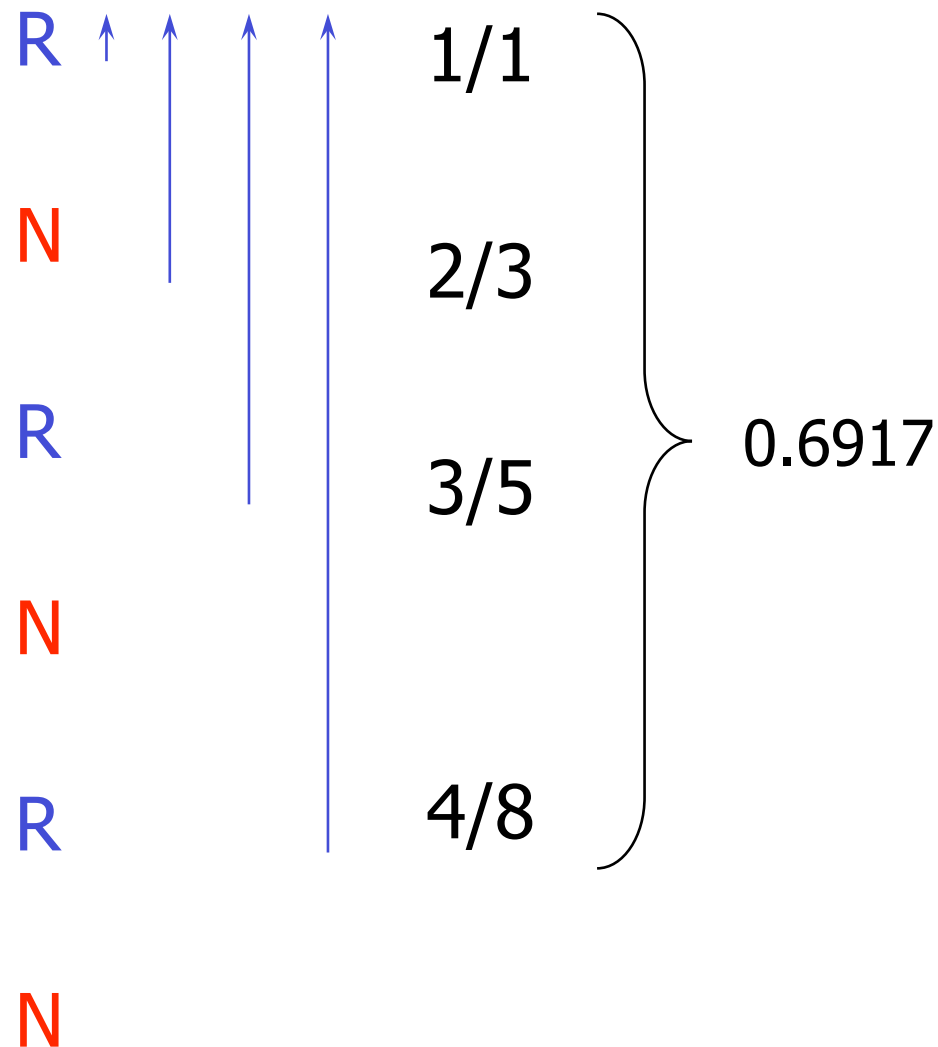
CombMNZ [Fox, Shaw, Lee, et al.]

- Normalize scores: $[0,1]$.
- For each doc:
 - sum relevance scores given to it by each system (use 0 if unretrieved), and
 - multiply by number of systems that retrieved it (MNZ).
- Rank documents by score.

How well do they perform?

- Need *performance metric*.
- Need *benchmark data*.

Metric: Average Precision



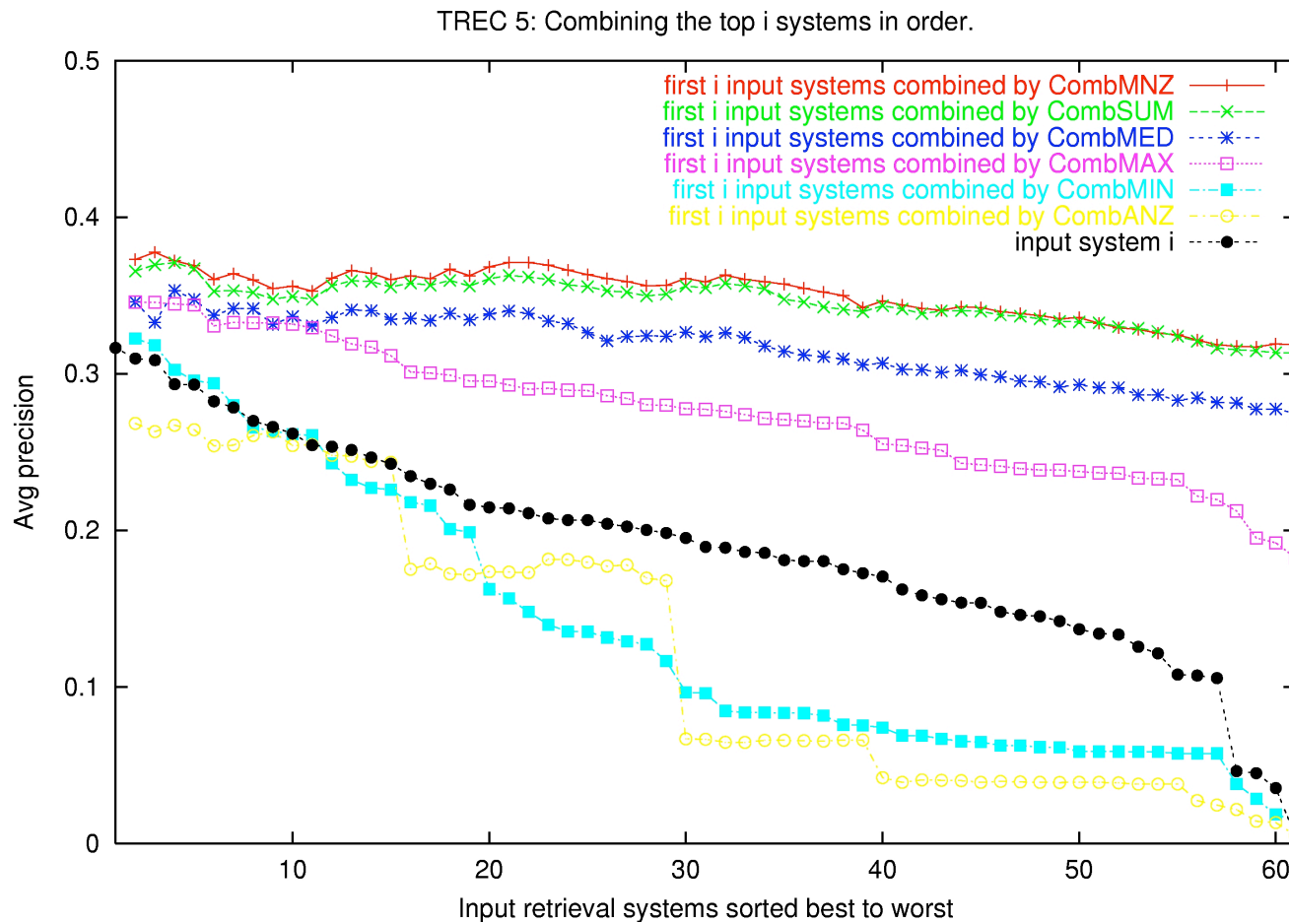
Benchmark Data: TREC

- Annual *Text Retrieval Conference*.
- Millions of documents (AP, NYT, etc.)
- 50 queries.
- Dozens of retrieval engines.
- Output lists available.
- Relevance judgments available.

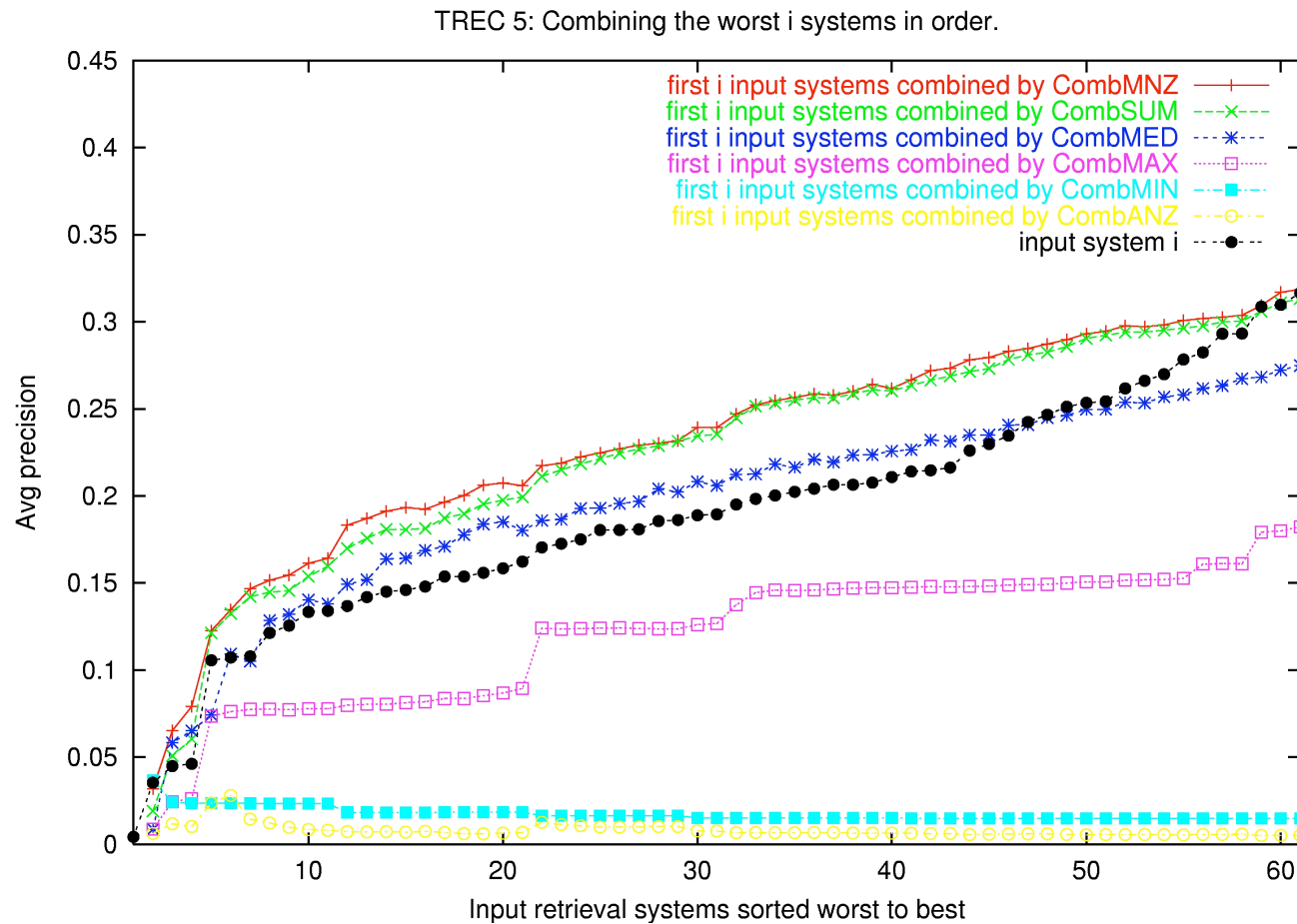
Data Sets

Data set	Number systems	Number queries	Number of docs
TREC3	40	50	1000
TREC5	61	50	1000
Vogt	10	10	1000
TREC9	105	50	1000

CombX on TREC5 Data



CombX on TREC5 Data, II

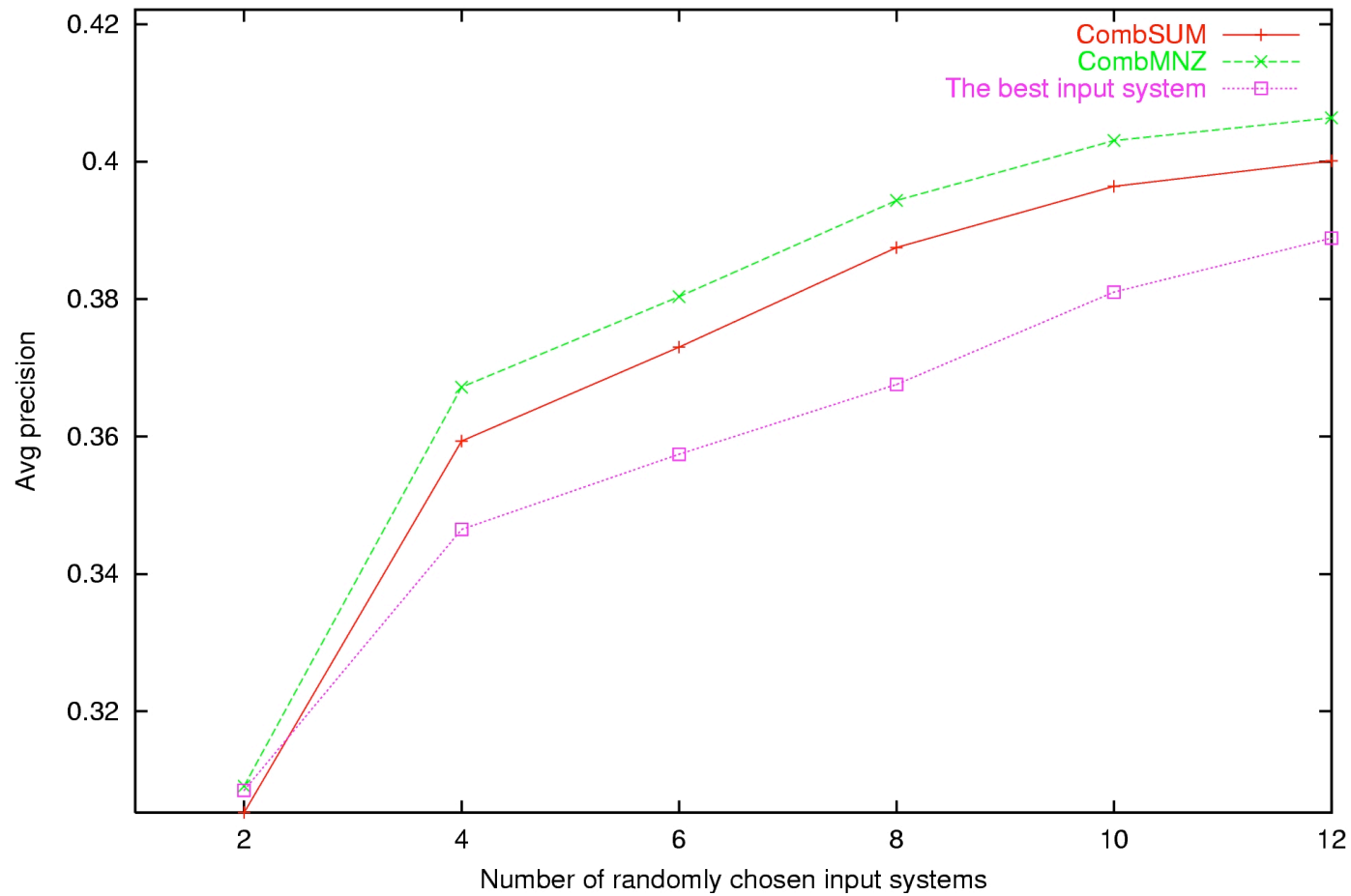


Experiments

- Randomly choose n input systems.
- For each query:
 - combine, trim, calculate avg precision.
- Calculate mean avg precision.
- Note best input system.
- Repeat (statistical significance).

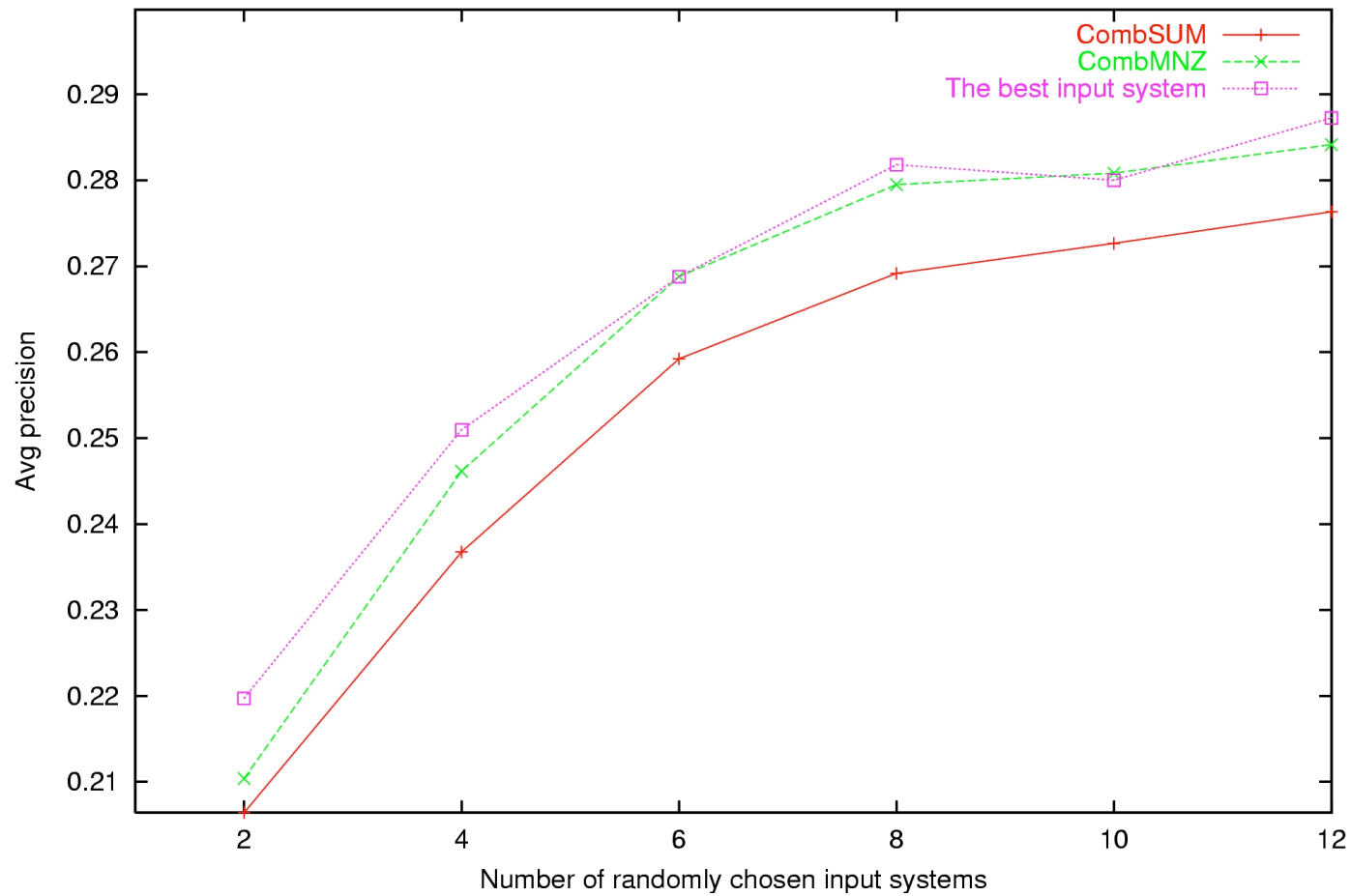
CombMNZ on TREC3

TREC 3: avg precision over 200 random sets of systems.



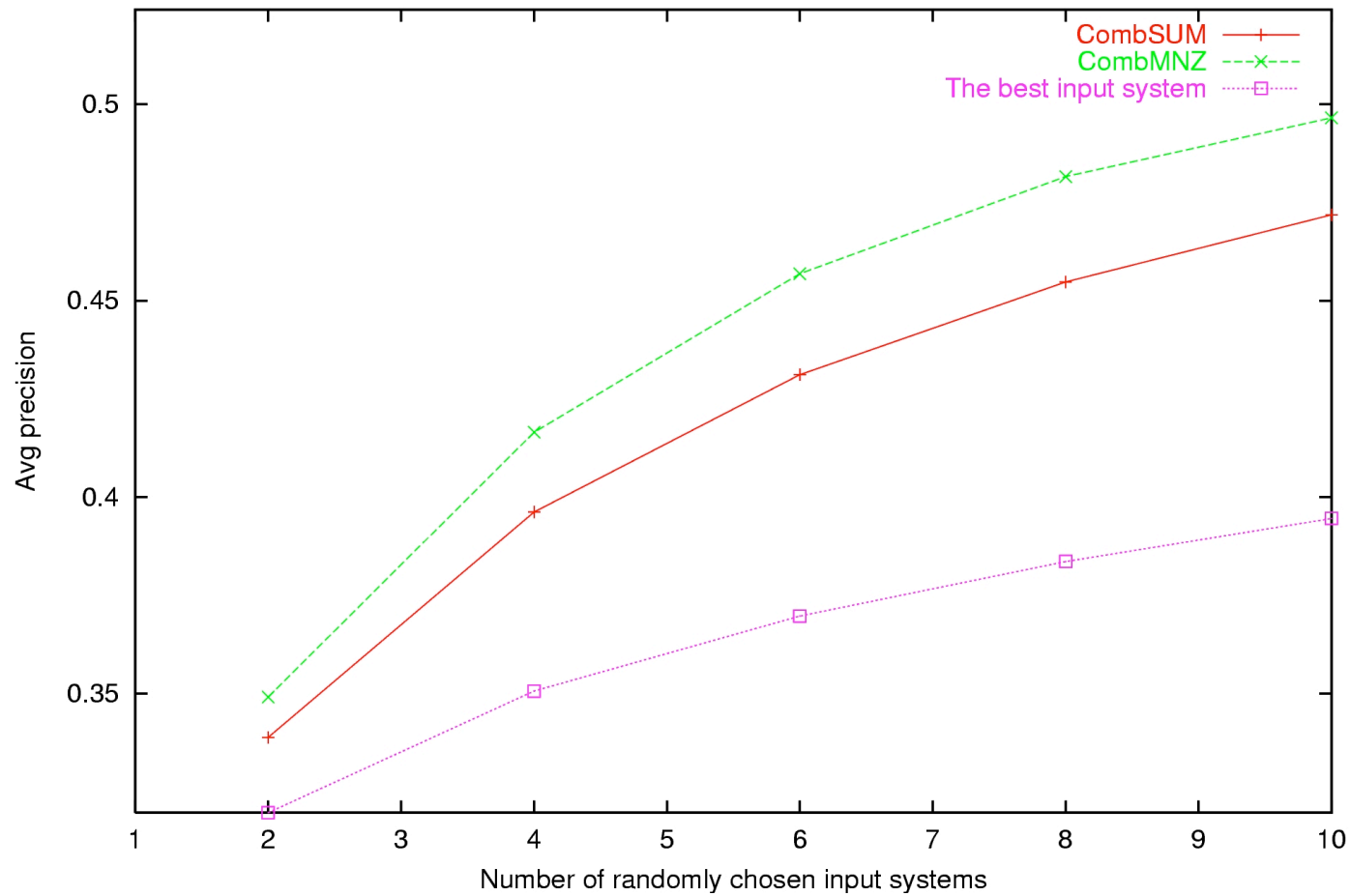
CombMNZ on TREC5

TREC 5: avg precision over 200 random sets of systems.



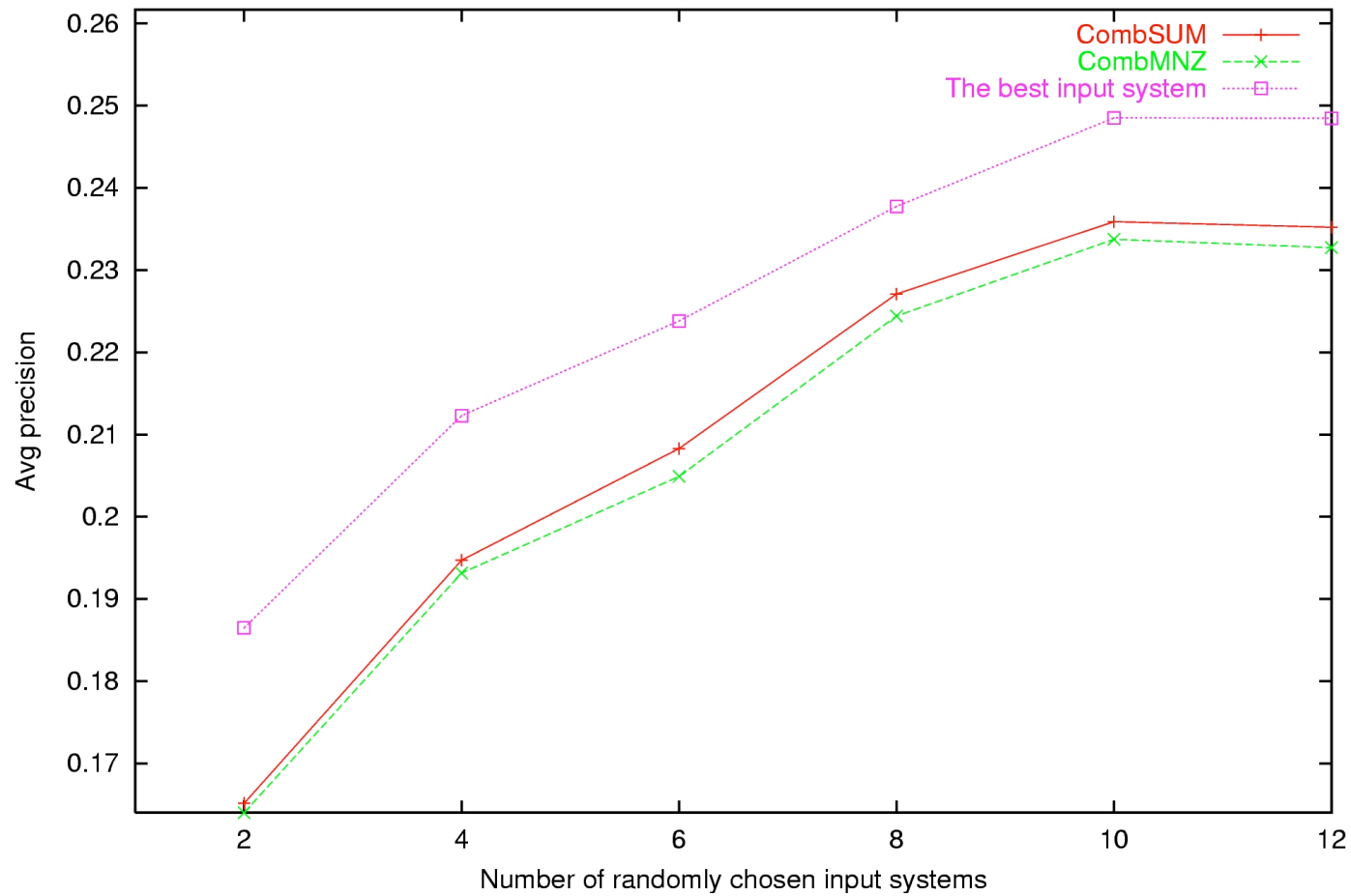
CombMNZ on Vogt

TREC 5 subset: avg precision over between 1 and 200 random sets of systems.



CombMNZ on TREC9

TREC 9: avg precision over 200 random sets of systems.



Metasearch via Voting

[Aslam, Montague]

- Analog to *election strategies*.
 - Requires only rank information.
 - No training required.

Classes of Metasearch Problems

	no training data	training data
ranks only	Borda, Condorcet, rCombMNZ	Bayes
relevance scores	CombMNZ	LC model

Election Strategies

- Plurality vote.
- Approval vote.
- Run-off.
- Preferential rankings:
 - instant run-off,
 - Borda count (positional),
 - Condorcet method (head-to-head).

Metasearch Analogy

- Documents are *candidates*.
- Systems are *voters* expressing preferential rankings among candidates.

Borda Count

- Consider an n candidate election.
- One method for choosing winner is the Borda count. [Borda, Saari]
 - For each voter i
 - Assign n points to top candidate.
 - Assign $n-1$ points to next candidate.
 - ...
 - Rank candidates according to point sum.

Election 2000: Florida

[NATIONAL](#) > [FLORIDA](#)

VOTER RESULTS IN FLORIDA | [EXIT POLLS](#) | [HOUSE AND SENATE](#)

Last updated: 12:32 a.m. EST, 12/14 | [refresh this page](#) | [print this page](#)

FLORIDA VOTE COUNT

	Nov. 7	Recount	Certified	12/8 Ruling
Bush <input checked="" type="checkbox"/>	1,725	930	537	193

Source: Associated Press

25 electoral votes at stake

PRESIDENT DEC. 13

100% of precincts

Candidates	Votes	Vote %	States Won	EV
R Bush <input checked="" type="checkbox"/>	2,909,176	49 %	29	0
D Gore	2,907,451	49 %	20	0
G Nader	96,837	2 %	0	0
I Browne	18,856	0 %	0	0
RF Buchanan	17,356	0 %	0	0
I Phillips	4,280	0 %	0	0
I Hagelin	2,287	0 %	0	0

winner declared

[exit polls](#)

results as of 5:46 p.m. EST

Borda Count: Election 2000

- Ideological order: Nader, Gore, Bush.
 - Ideological voting:
 - Bush voter: Bush, Gore, Nader.
 - Nader voter: Nader, Gore, Bush.
 - Gore voter:
 - Gore, Bush, Nader.
 - Gore, Nader, Bush.
- } 50/50, 100/0

Election 2000: Ideological Florida Voting

	Gore	Bush	Nader
50/50	14,734,379	13,185,542	7,560,864
100/0	14,734,379	14,639,267	6,107,138

Gore Wins

Borda Count: Election 2000

- Ideological order: Nader, Gore, Bush.
- Manipulative voting:
 - Bush voter: Bush, Nader, Gore.
 - Gore voter: Gore, Nader, Bush.
 - Nader voter: Nader, Gore, Bush.

Election 2000: Manipulative Florida Voting

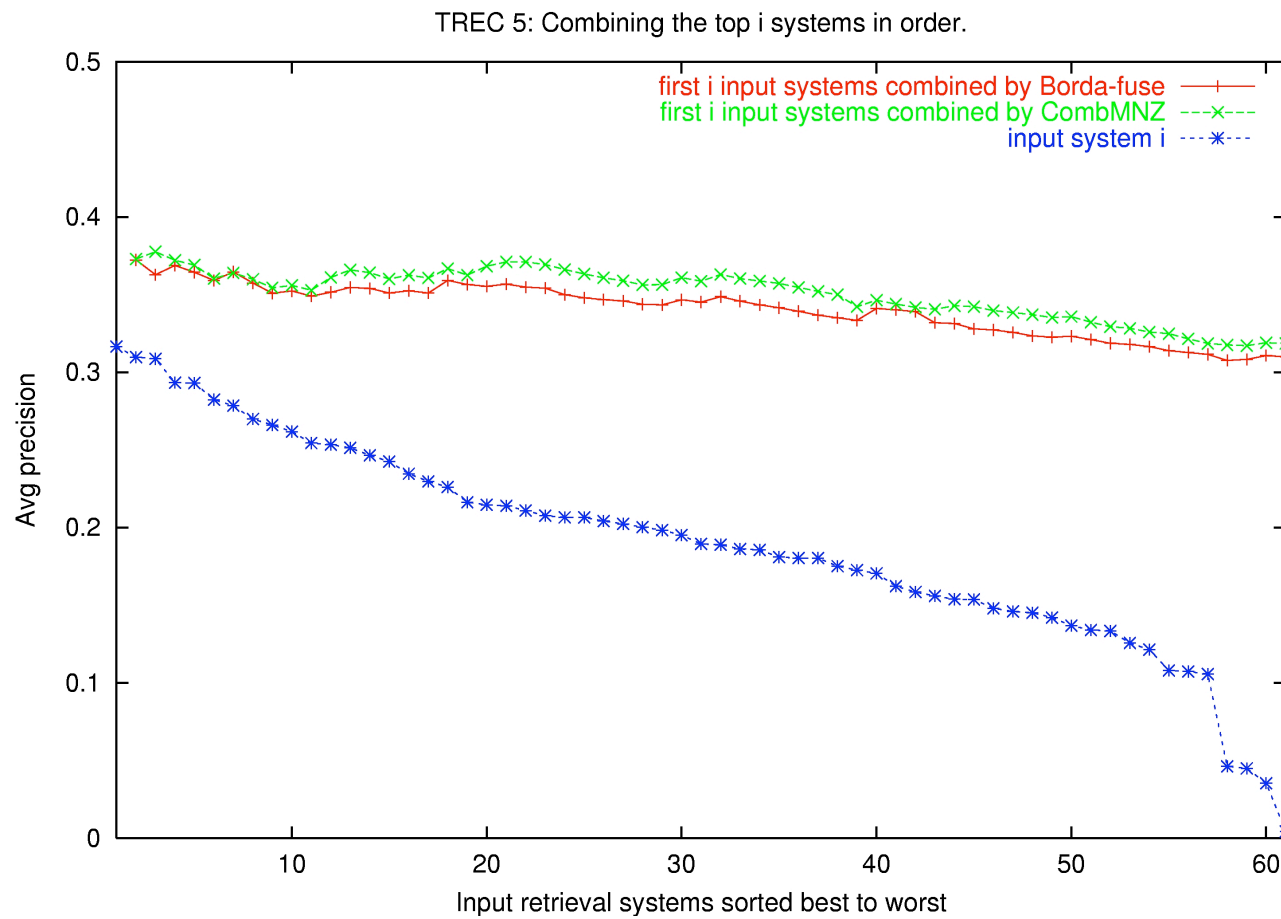
Gore	Bush	Nader
11,825,203	11,731,816	11,923,765

Nader Wins

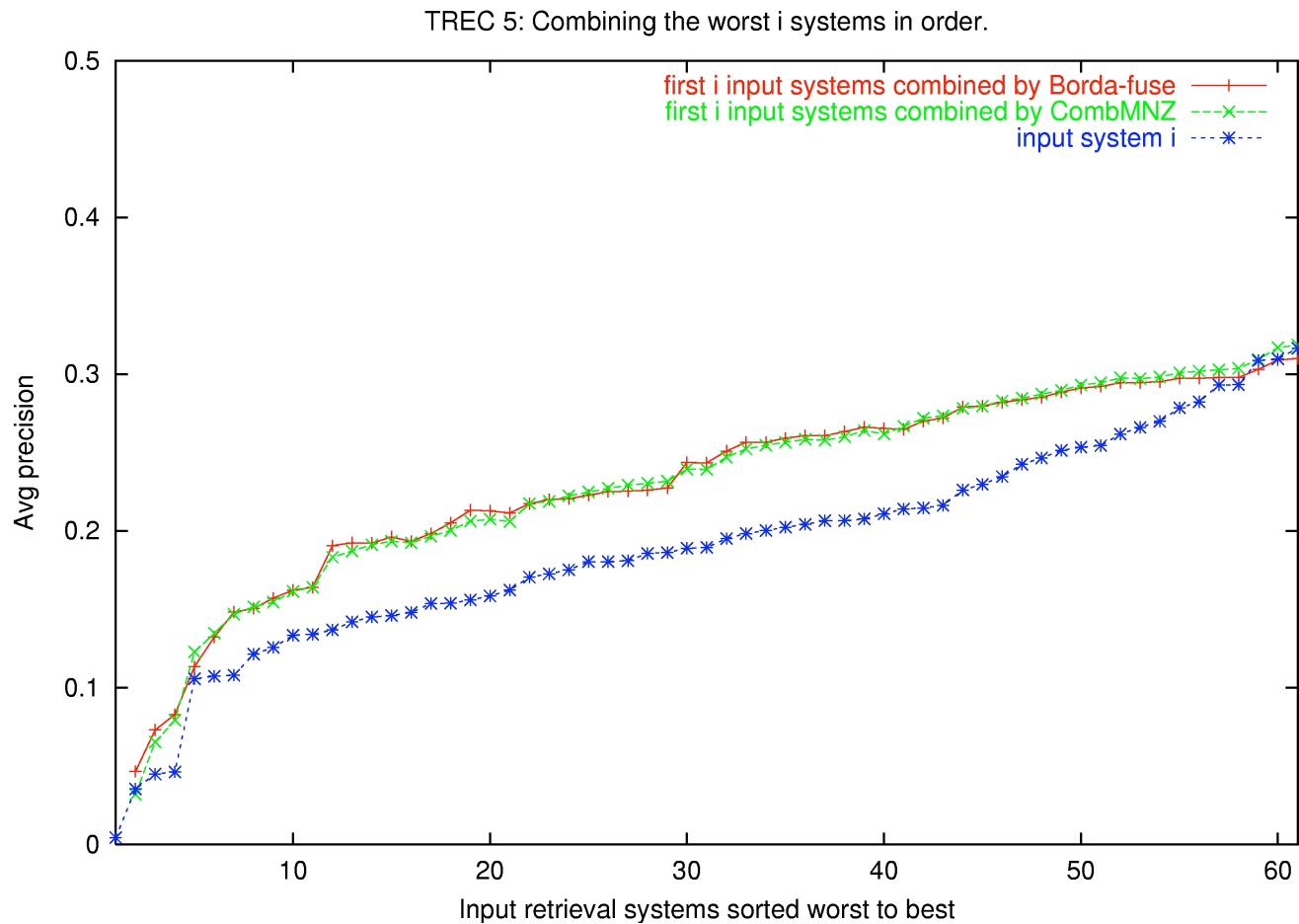
Metasearch via Borda Counts

- Metasearch analogy:
 - Documents are *candidates*.
 - Systems are *voters* providing preferential rankings.
- Issues:
 - Systems may rank different document sets.
 - How to deal with unranked documents?

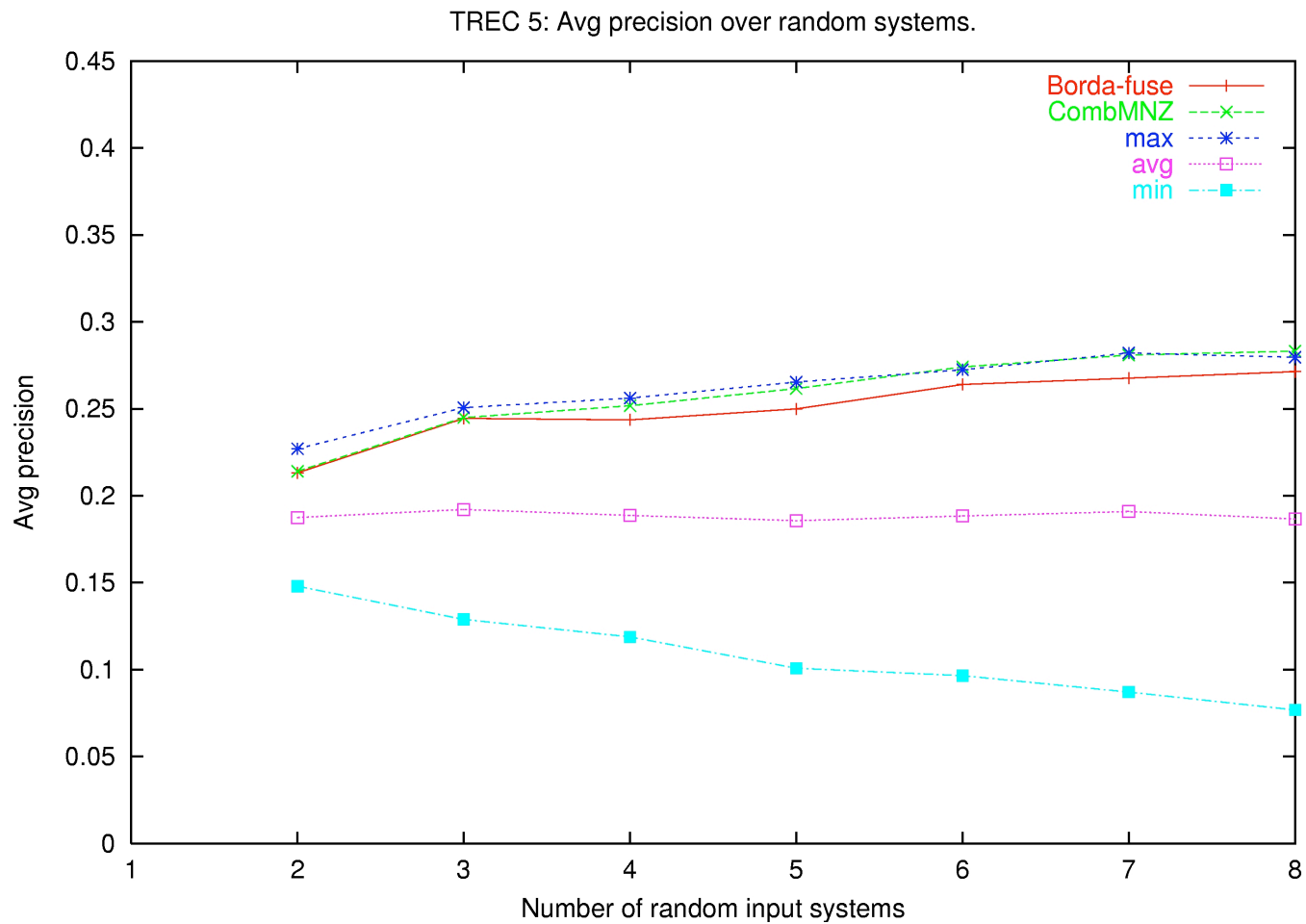
Borda on TREC5 Data, I



Borda on TREC5 Data, II



Borda on TREC5 Data, III



Condorcet Voting

- Each ballot ranks all candidates.
- Simulate head-to-head run-off between each pair of candidates.
- Condorcet winner: candidate that beats all other candidates, head-to-head.

Election 2000: Florida

[NATIONAL](#) > [FLORIDA](#)

VOTER RESULTS IN FLORIDA | [EXIT POLLS](#) | [HOUSE AND SENATE](#)

Last updated: 12:32 a.m. EST, 12/14 | [refresh this page](#) | [print this page](#)

FLORIDA VOTE COUNT

	Nov. 7	Recount	Certified	12/8 Ruling
Bush <input checked="" type="checkbox"/>	1,725	930	537	193

Source: Associated Press

25 electoral votes at stake

PRESIDENT DEC. 13

100% of precincts

Candidates	Votes	Vote %	States Won	EV
R Bush <input checked="" type="checkbox"/>	2,909,176	49 %	29	0
D Gore	2,907,451	49 %	20	0
G Nader	96,837	2 %	0	0
I Browne	18,856	0 %	0	0
RF Buchanan	17,356	0 %	0	0
I Phillips	4,280	0 %	0	0
I Hagelin	2,287	0 %	0	0

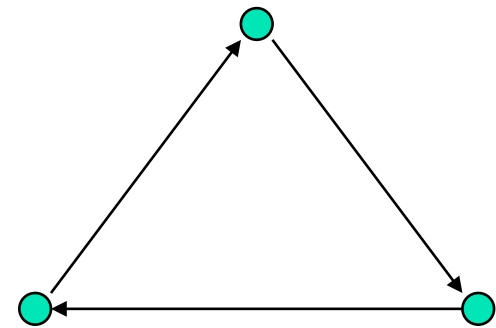
winner declared

[exit polls](#)

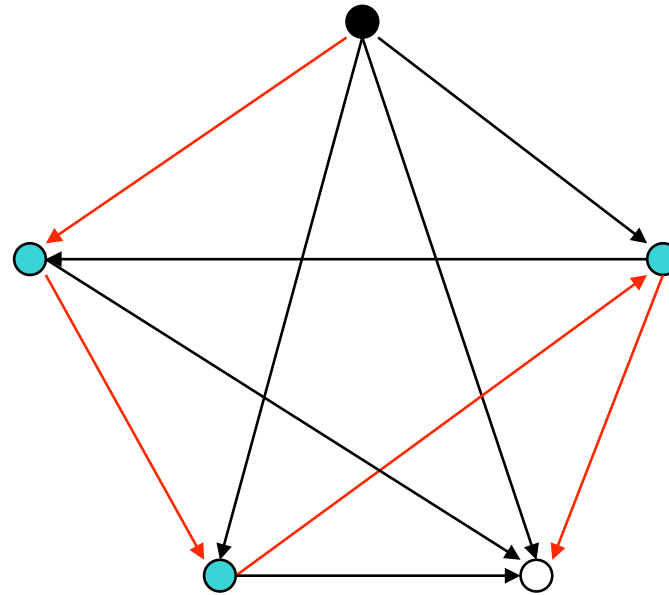
results as of 5:46 p.m. EST

Condorcet Paradox

- Voter 1: A, B, C
- Voter 2: B, C, A
- Voter 3: C, A, B
- Cyclic preferences: cycle in Condorcet graph.
- Condorcet consistent path: Hamiltonian.
- For metasearch: any CC path will do.



Condorcet Consistent Path

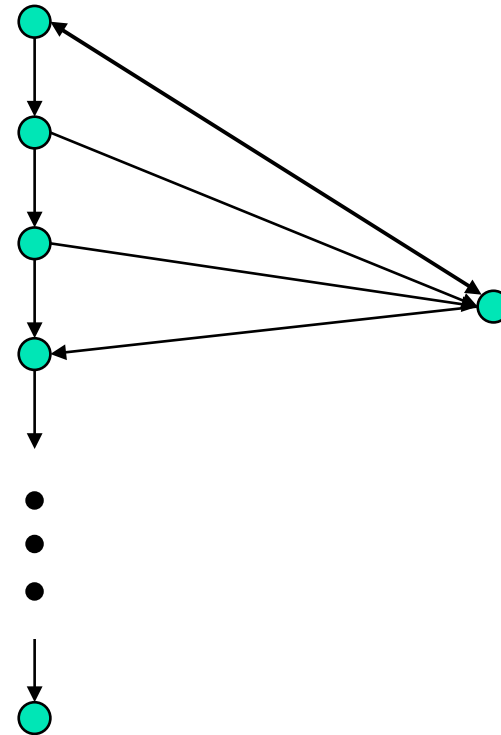


Hamiltonian Path Proof

Base Case:



Inductive Step:

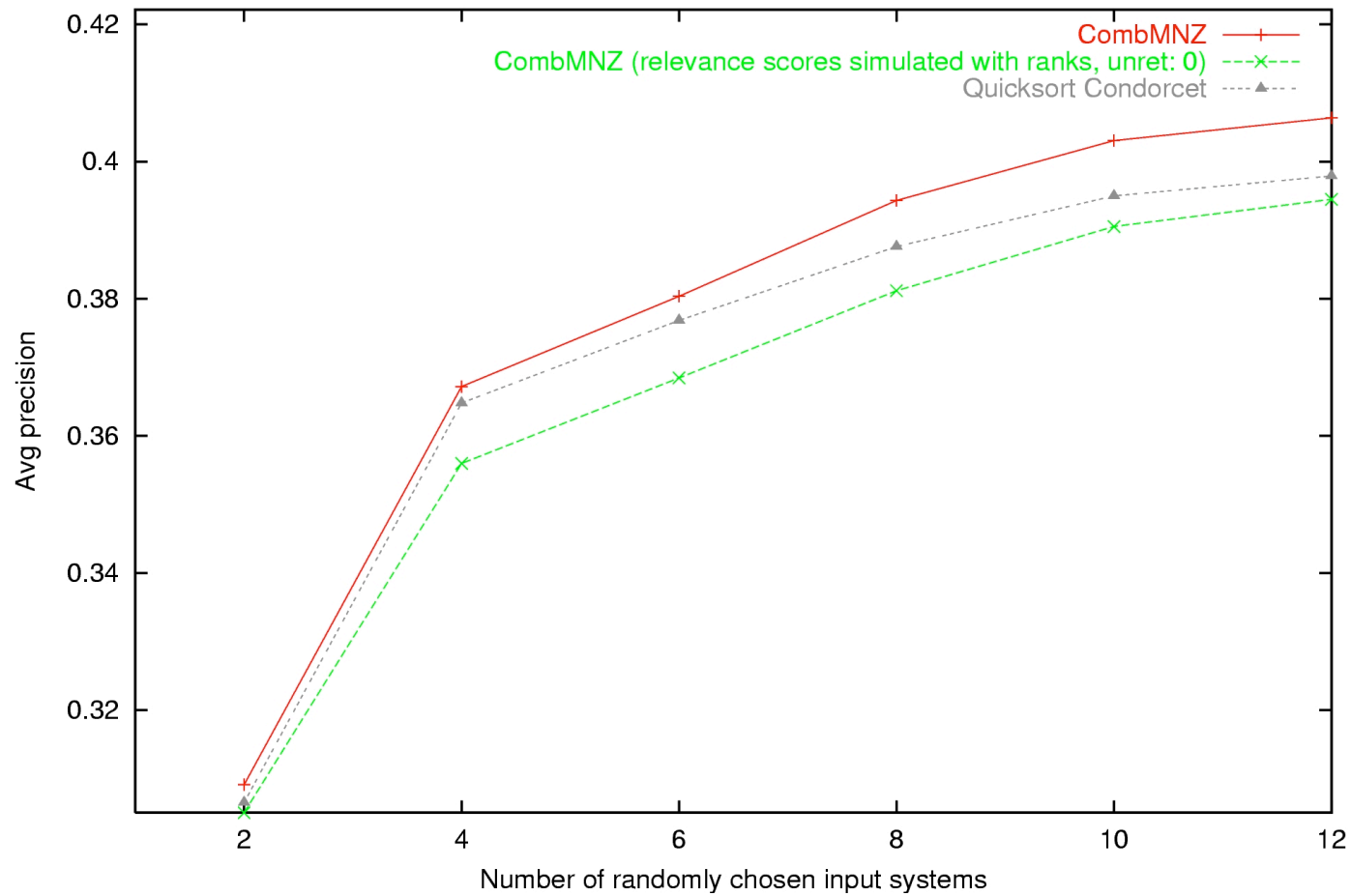


Condorcet-fuse: Sorting

- Insertion-sort suggested by proof.
- Quicksort too; $O(n \log n)$ comparisons.
 - n documents.
- Each comparison: $O(m)$.
 - m input systems.
- Total: $O(m n \log n)$.
- *Need not compute entire graph.*

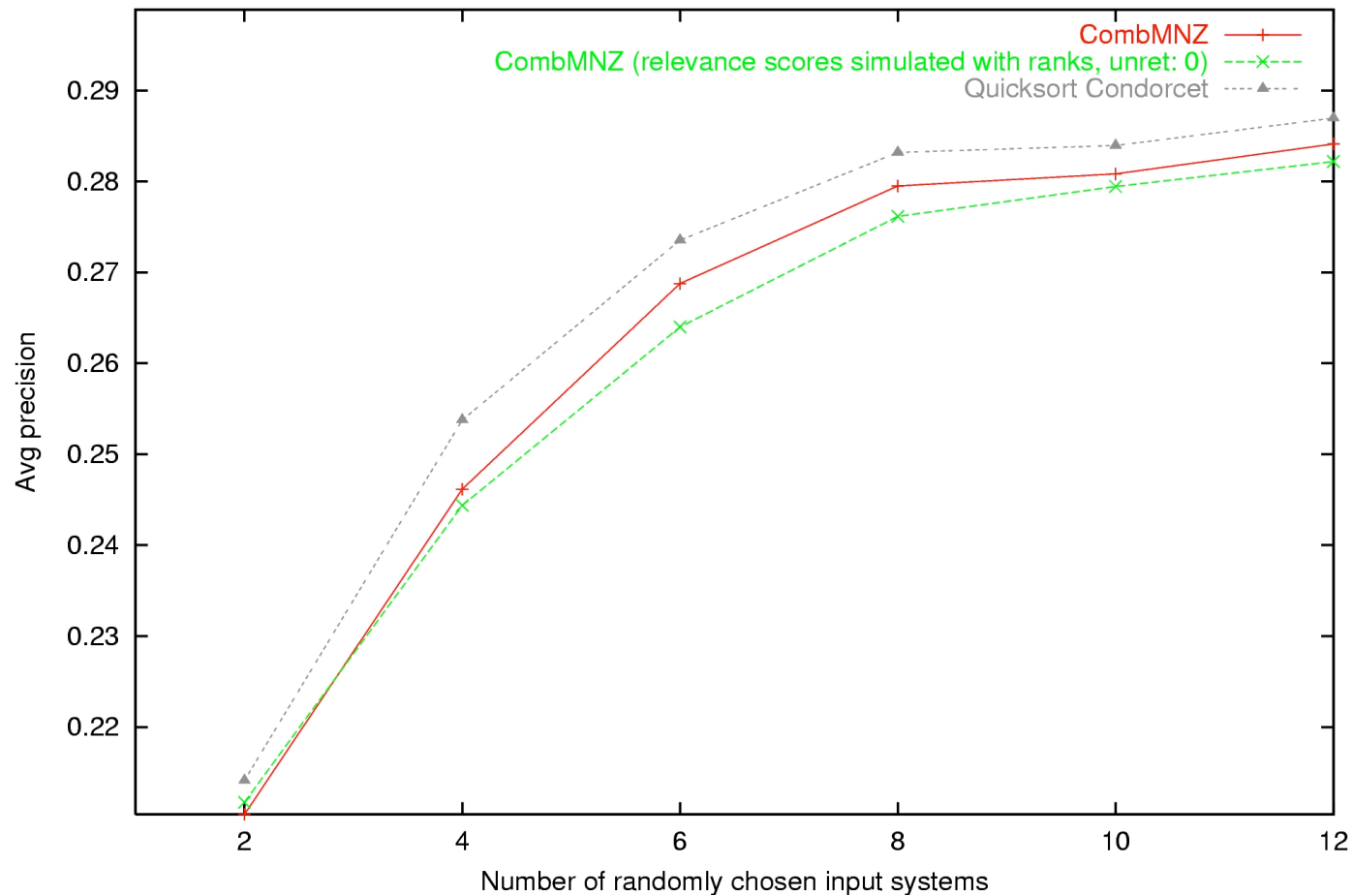
Condorcet-fuse on TREC3

TREC 3: avg precision over 200 random sets of systems.



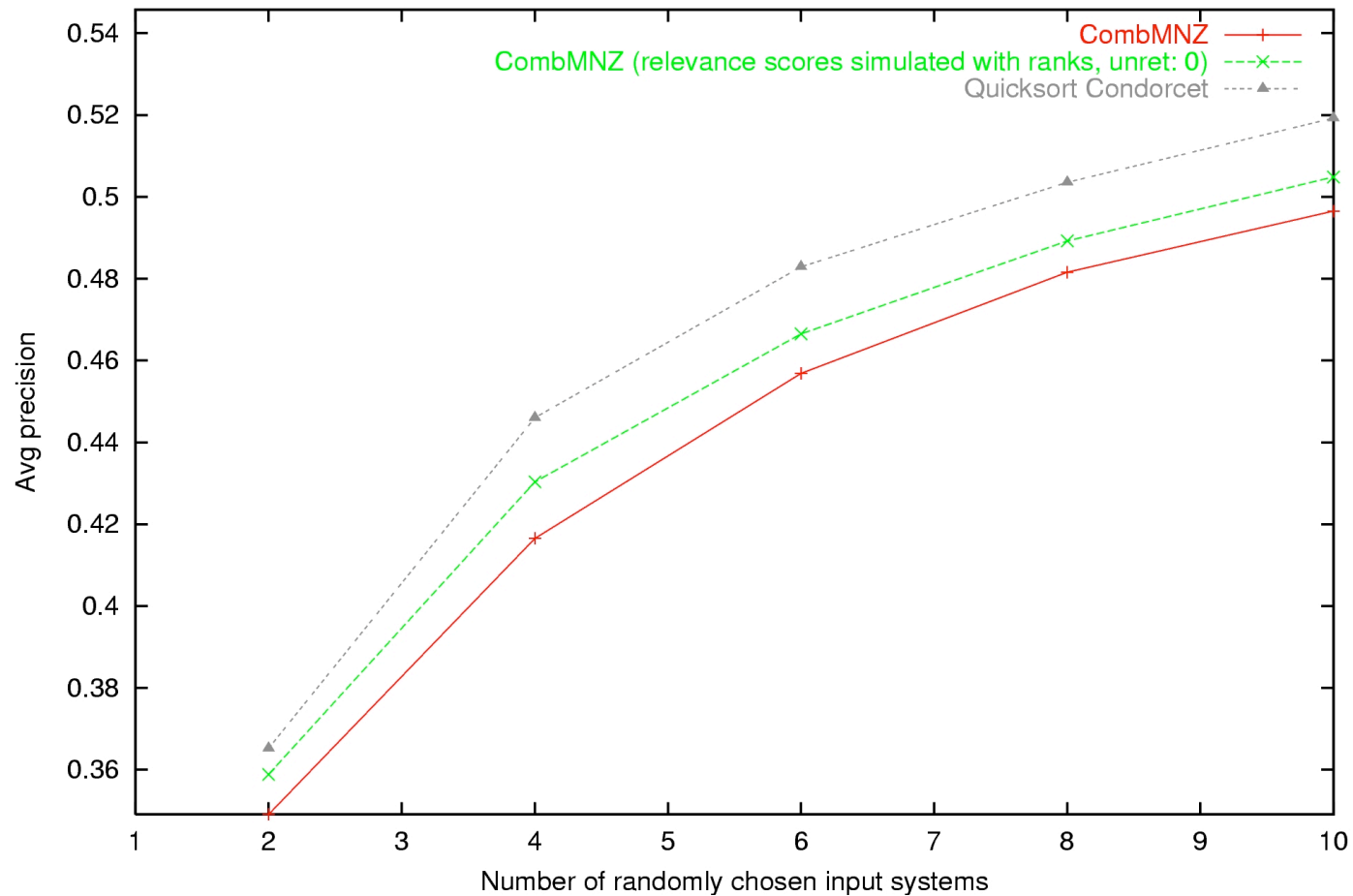
Condorcet-fuse on TREC5

TREC 5: avg precision over 200 random sets of systems.



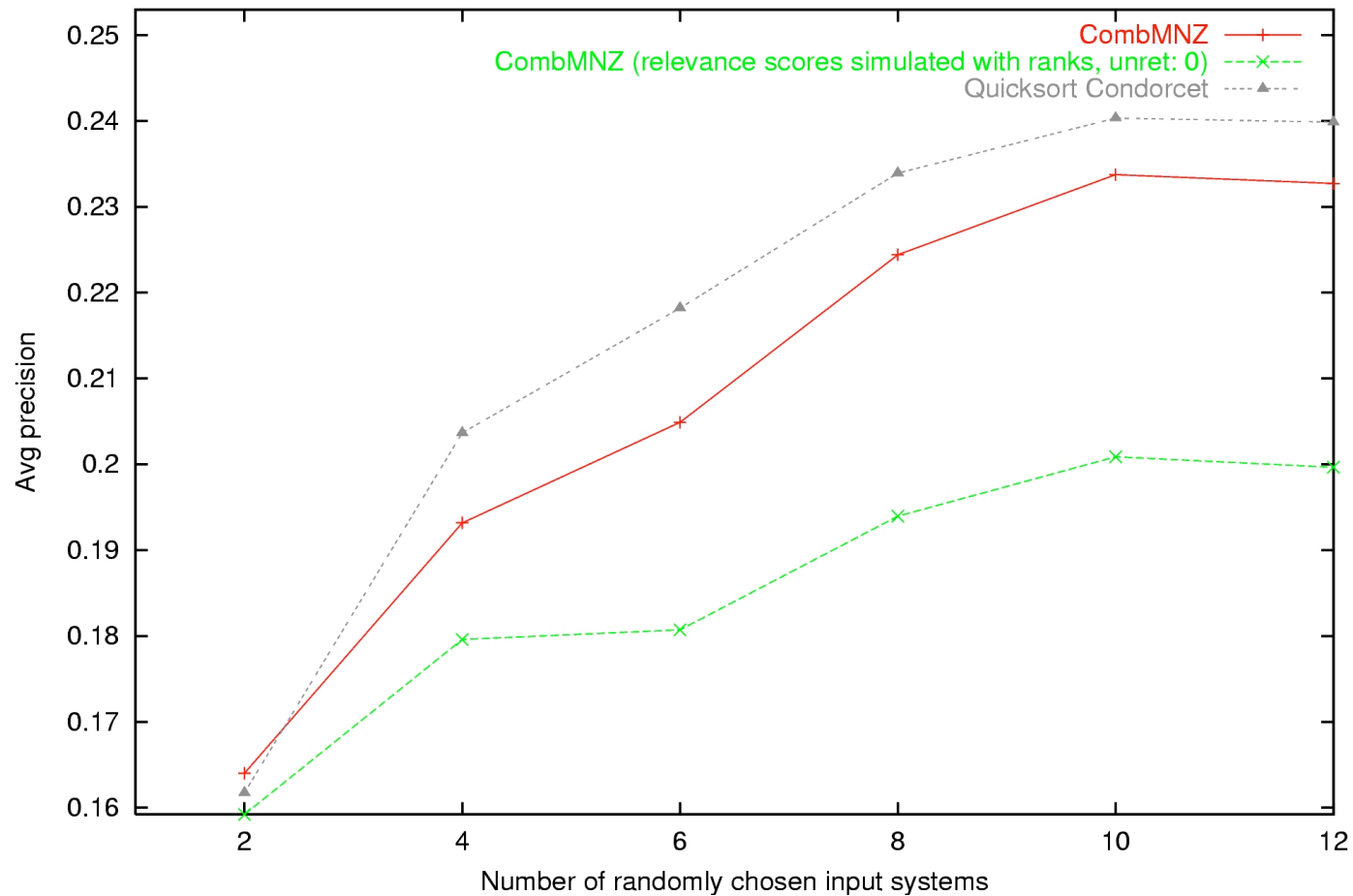
Condorcet-fuse on Vogt

TREC 5 subset: avg precision over between 1 and 200 random sets of systems.



Condorcet-fuse on TREC9

TREC 9: avg precision over 200 random sets of systems.



Outline

- ✓ Introduce problem
- ✓ Characterize problem
- ✓ Survey techniques
- Upper bounds for metasearch

Upper Bounds on Metasearch

- How good can metasearch be?
- Are there fundamental limits that methods are approaching?

Upper Bounds on Metasearch

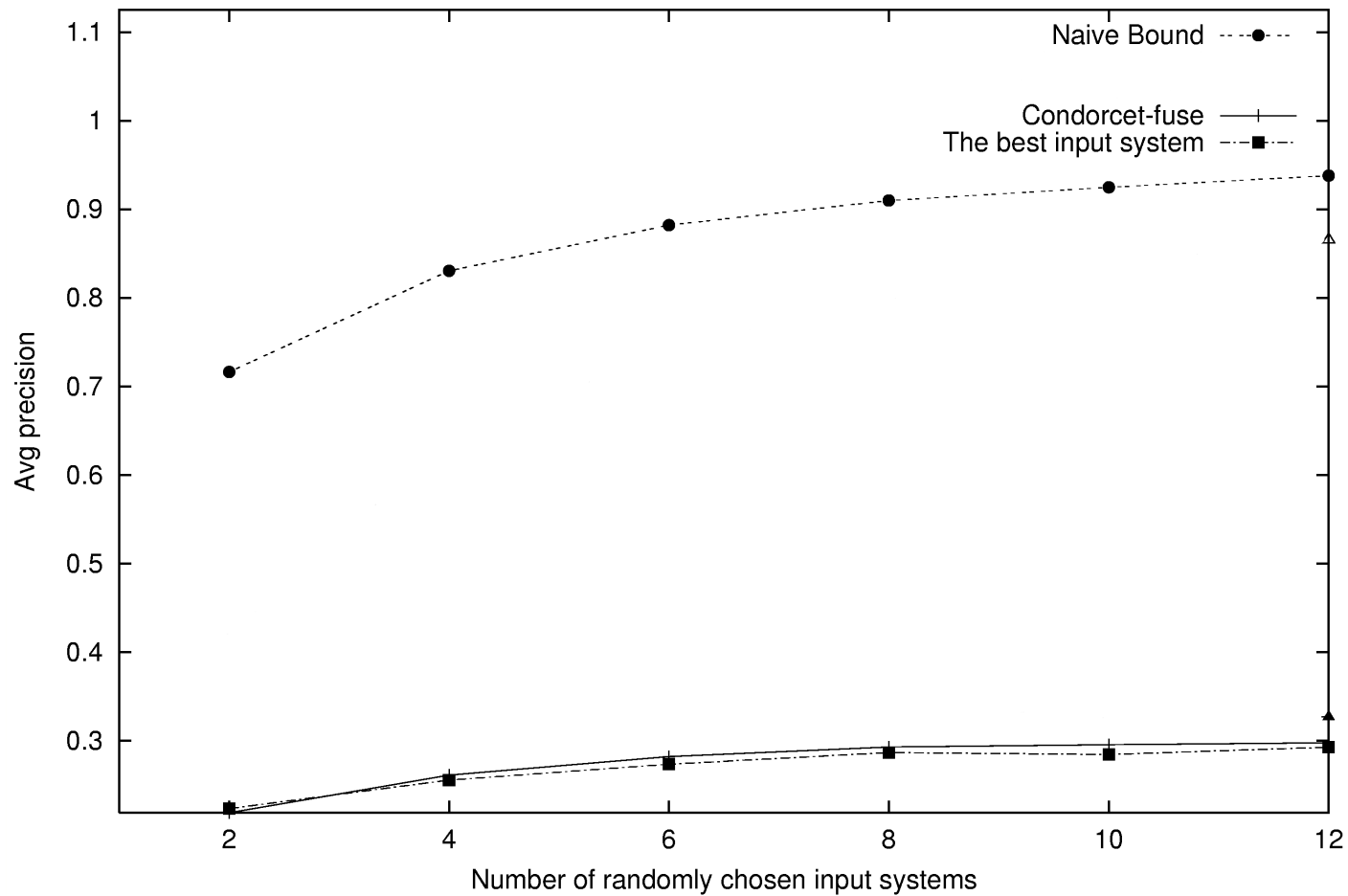
- Constrained oracle model:
 - omniscient metasearch oracle,
 - constraints placed on oracle that any reasonable metasearch technique must obey.
- What are “reasonable” constraints?

Naïve Constraint

- *Naïve* constraint:
 - Oracle may only return docs from underlying lists.
 - Oracle may return these docs in any order.
 - Omniscient oracle will return relevant docs above irrelevant docs.

TREC5: Naïve Bound

TREC 5: avg precision over 200 random sets of systems.

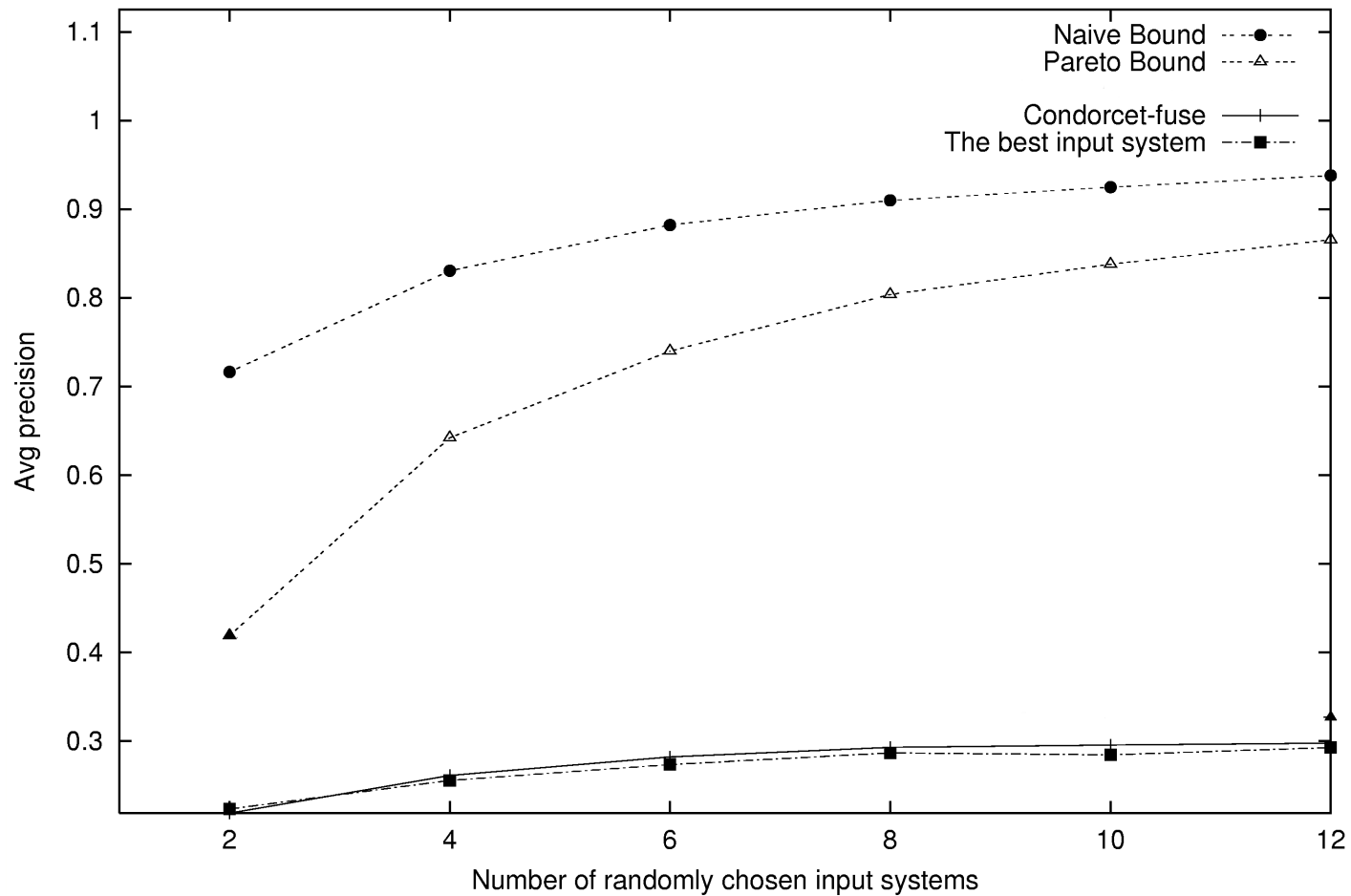


Pareto Constraint

- *Pareto* constraint:
 - Oracle may only return docs from underlying lists.
 - Oracle must respect *unanimous* will of underlying systems.
 - Omniscient oracle will return relevant docs above irrelevant docs, subject to the above constraint.

TREC5: Pareto Bound

TREC 5: avg precision over 200 random sets of systems.

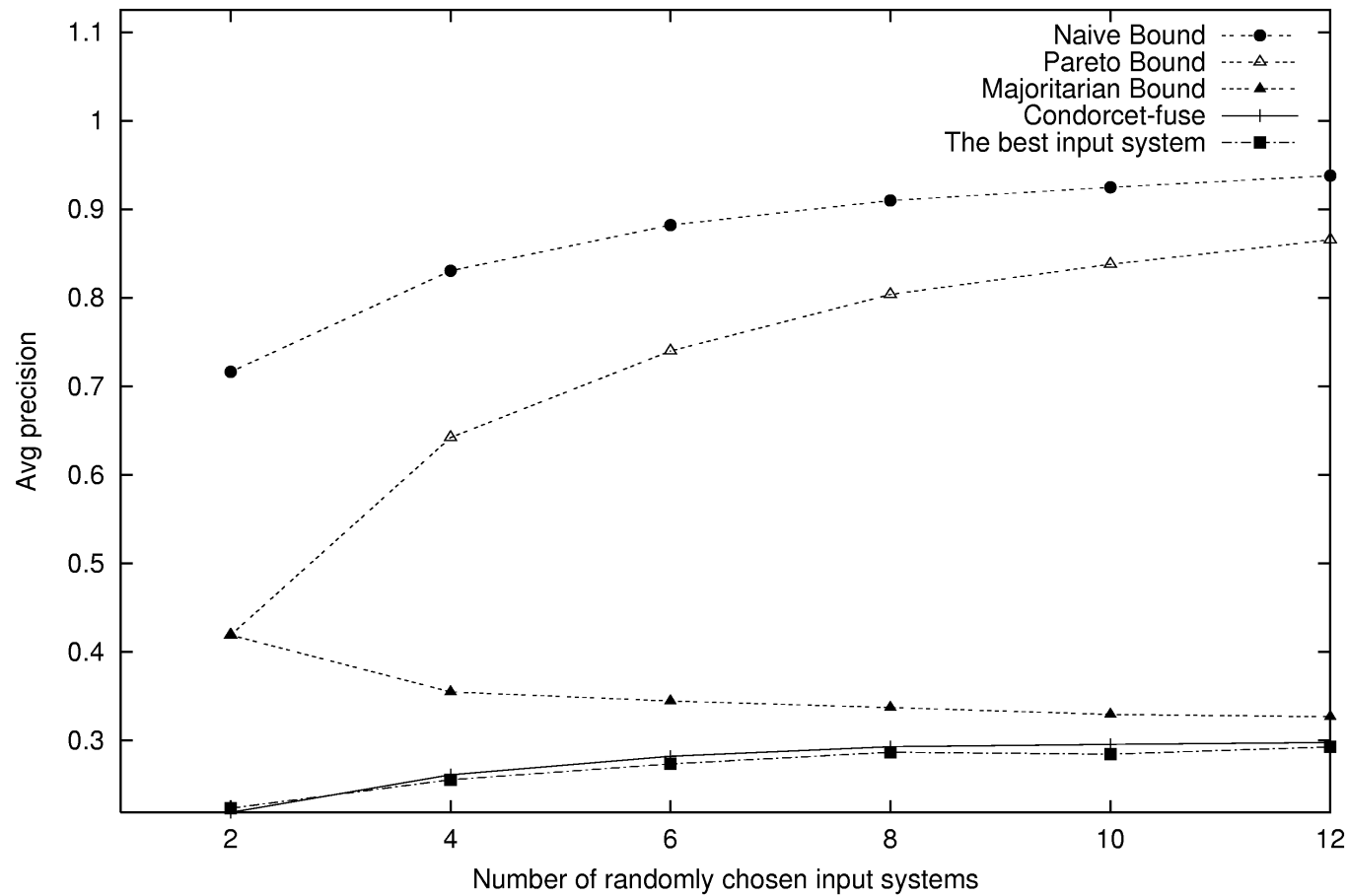


Majoritarian Constraint

- *Majoritarian* constraint:
 - Oracle may only return docs from underlying lists.
 - Oracle must respect *majority* will of underlying systems.
 - Omniscient oracle will return relevant docs above irrelevant docs and break cycles optimally, subject to the above constraint.

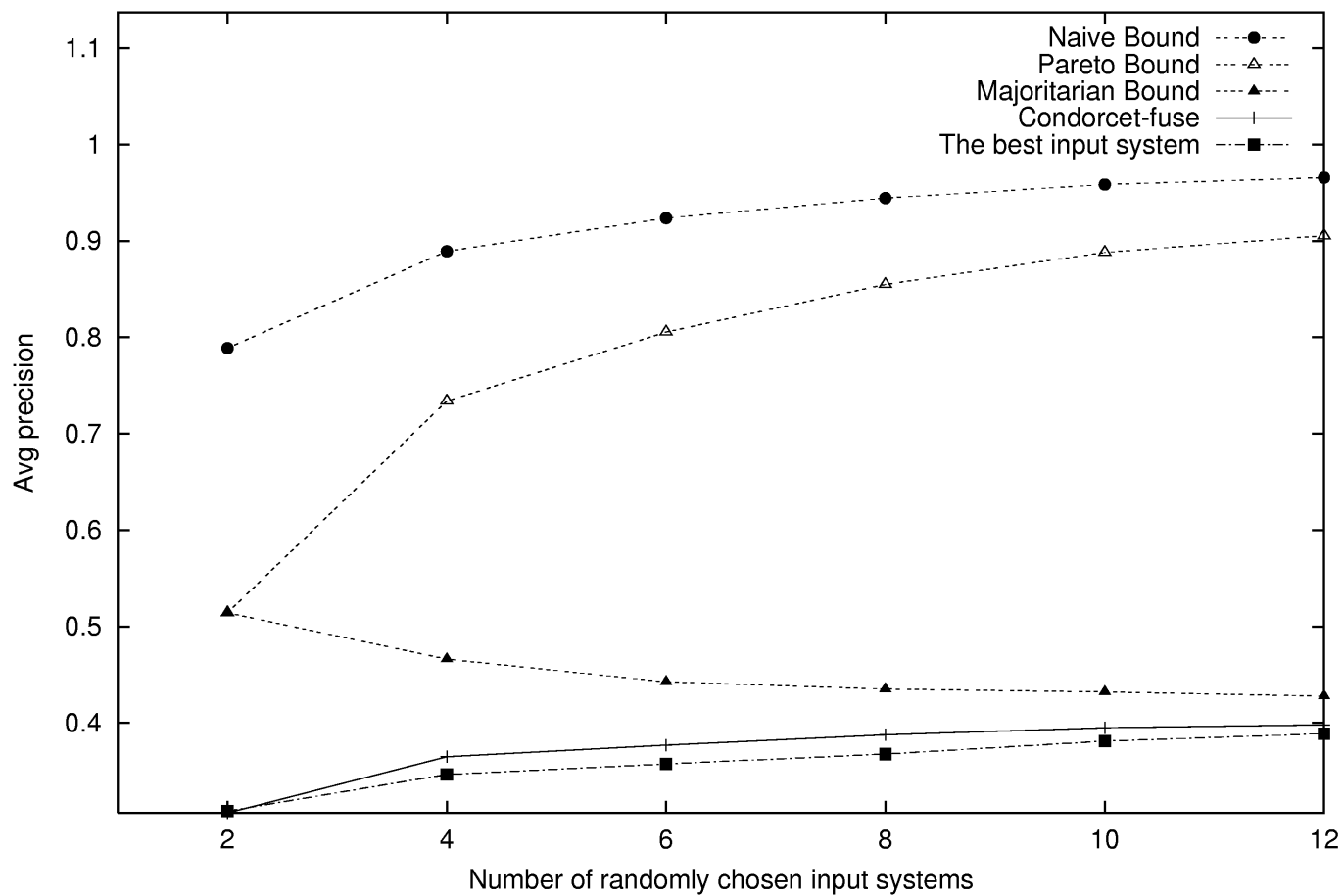
TREC5: Majoritarian Bound

TREC 5: avg precision over 200 random sets of systems.



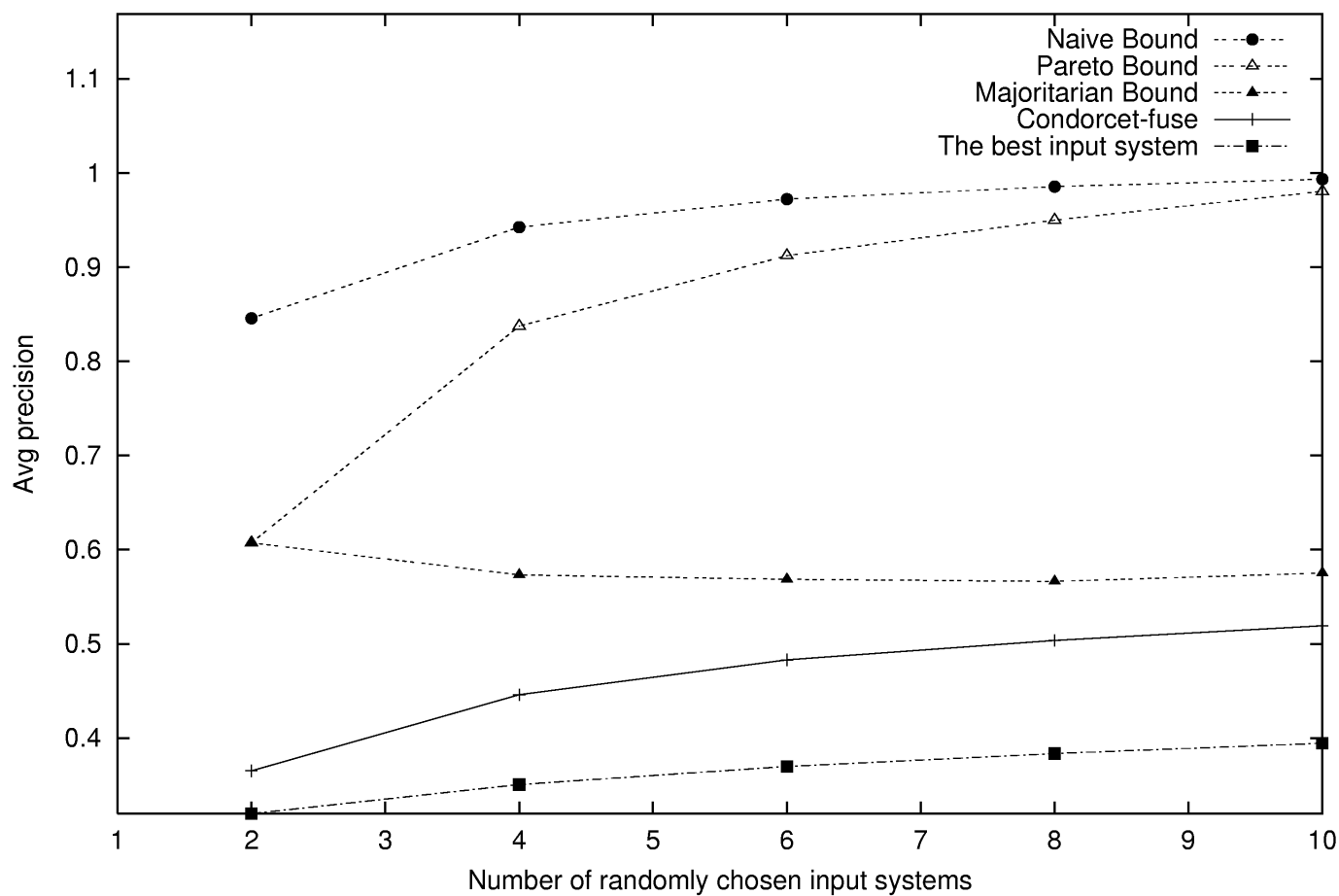
Upper Bounds: TREC3

TREC 3: avg precision over 200 random sets of systems.



Upper Bounds: Vogt

TREC 5 subset: avg precision over between 1 and 200 random sets of systems.



Upper Bounds: TREC9

TREC 9: avg precision over 200 random sets of systems.

