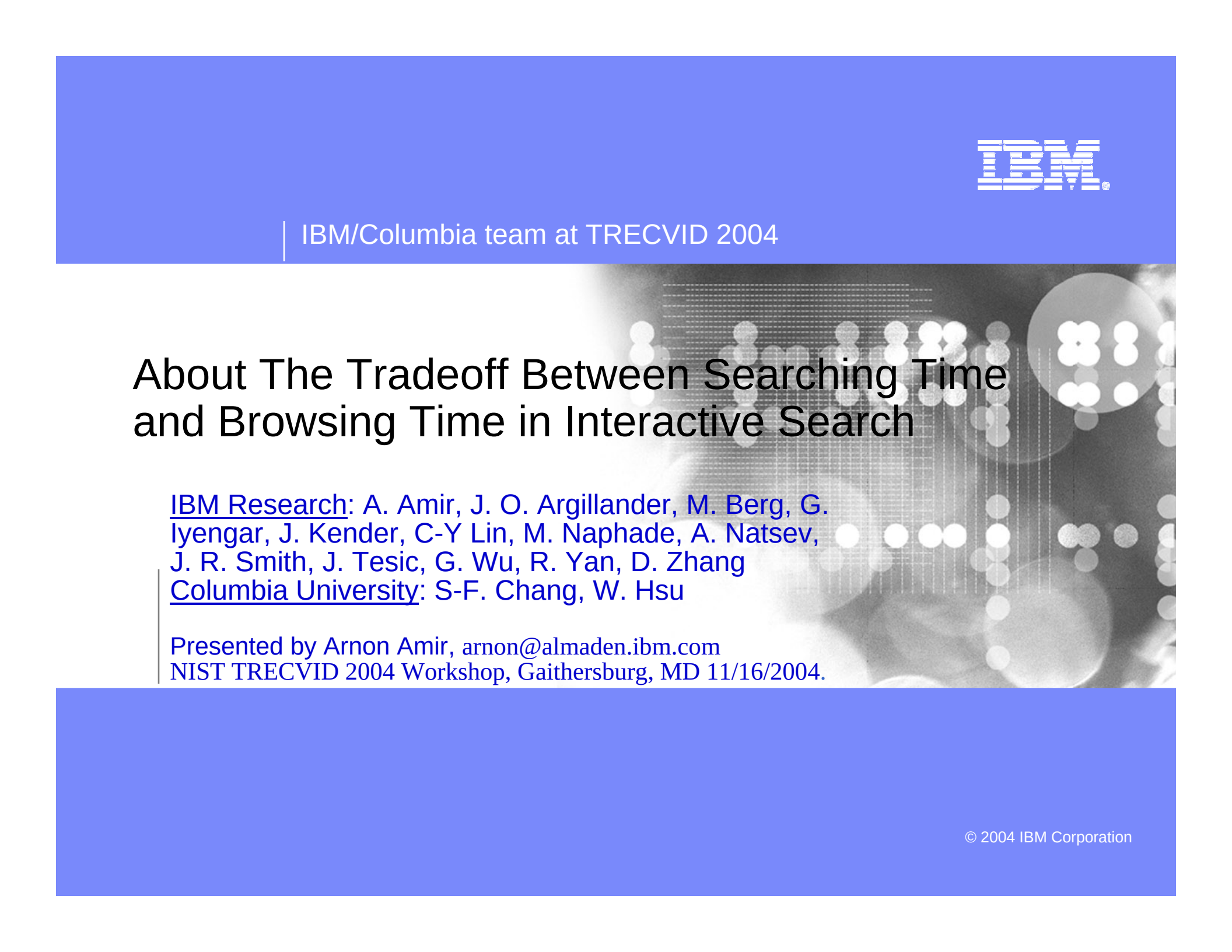


IBM/Columbia team at TRECVID 2004

About The Tradeoff Between Searching Time and Browsing Time in Interactive Search

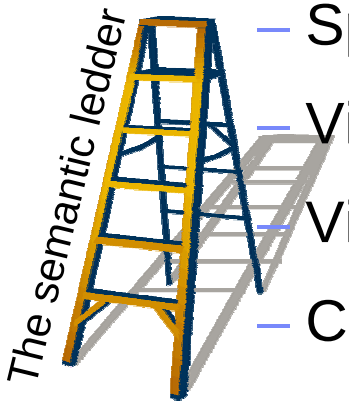
IBM Research: A. Amir, J. O. Argillander, M. Berg, G. Iyengar, J. Kender, C-Y Lin, M. Naphade, A. Natsev, J. R. Smith, J. Tesic, G. Wu, R. Yan, D. Zhang
Columbia University: S-F. Chang, W. Hsu

Presented by Arnon Amir, arnon@almaden.ibm.com
NIST TRECVID 2004 Workshop, Gaithersburg, MD 11/16/2004.

© 2004 IBM Corporation

IBM Search Systems – indexed data

- **Multimodal query retrieval:**
  - Speech: combined ASR, CC and Phonetic
  - Video OCR
  - Visual concepts
  - CBIR: color histograms, color correlograms, texture.



We extend our thanks to

CLIPS-IMAG – reference shot list

CMU – Video OCR and time alignment of CC and ASR

LDC – Video data, closed captions

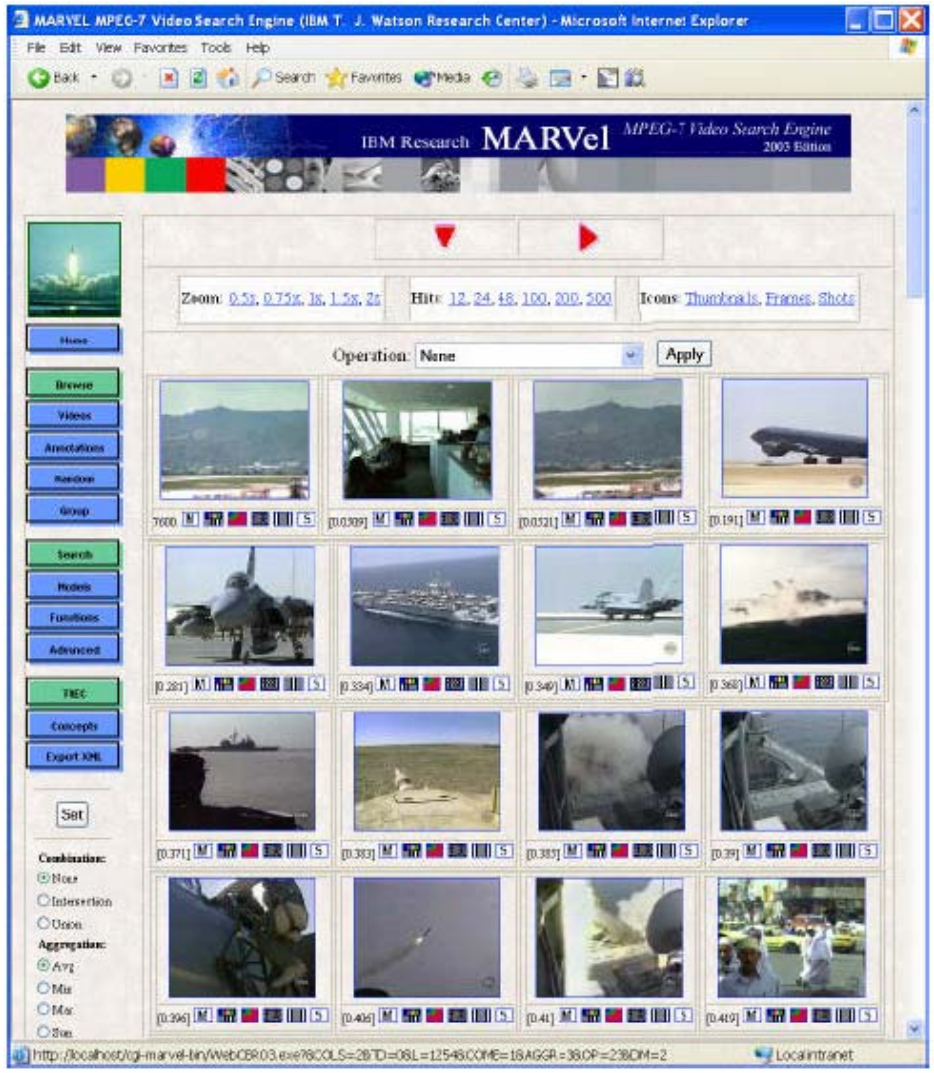
LIMSI – ASR data

NIST – TRECVID

© 2004 IBM Corporation

System 1: Marvel (“TJW”)

- Shot-based retrieval
- Speech, CC
- CBIR
- Concepts, models, filters
- Session
- Vector operations, fusion
- Shot aggregation
- Internet based GUI



© 2004 IBM Corporation

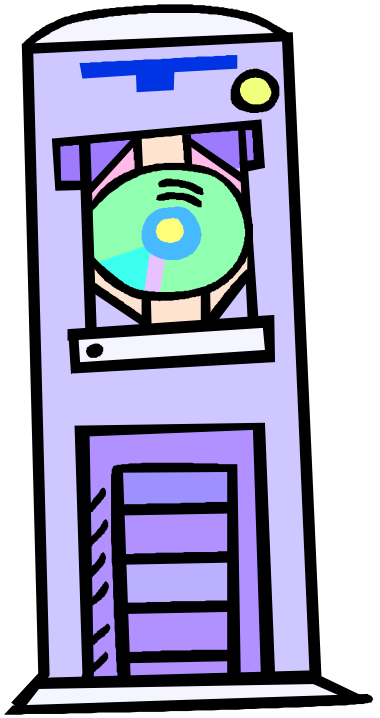
System 2: Multimedia Mining ("ARC")

- Video+time retrieval
- Speech, CC, phonetic
- CBIR
- Concepts, Video OCR
- Session
- Statistic ranking
- Shot aggregation
- Relevance Feedback
- Internet based GUI

© 2004 IBM Corporation

System 3: Automatic Search

- Shot-based retrieval
- Speech, CC
- CBIR
- Concepts
- Vector operations, fusion
- Automatic query formulation
  - No human in the loop, no GUI☺



© 2004 IBM Corporation

A Perspective on TRECVID Search

- What Manual search is?
- What Interactive search is?
- What is the importance of browsing?
- What is the importance of shots elevation?

© 2004 IBM Corporation

Manual Search

- Input: Multimodal topic, textual + image/video samples
- *Task: Form the best MM query for the given topic*
- Measure system+human performance
  - + Human ability to form a good query
  - + System discrimination
  - + System generalization
- Final search query is test-set independent, scalable

© 2004 IBM Corporation

Interactive Search

- Input: Multimodal topic, textual + image/video samples
- *Task: Get me as many hits as possible, now!*
- Measure system+human performance
  - + Human ability to form a good query
  - + Human browsing skills
  - + System search capabilities
  - + System browsing capabilities
- Final search result is test-set dependent, not scalable

© 2004 IBM Corporation

Automatic Search

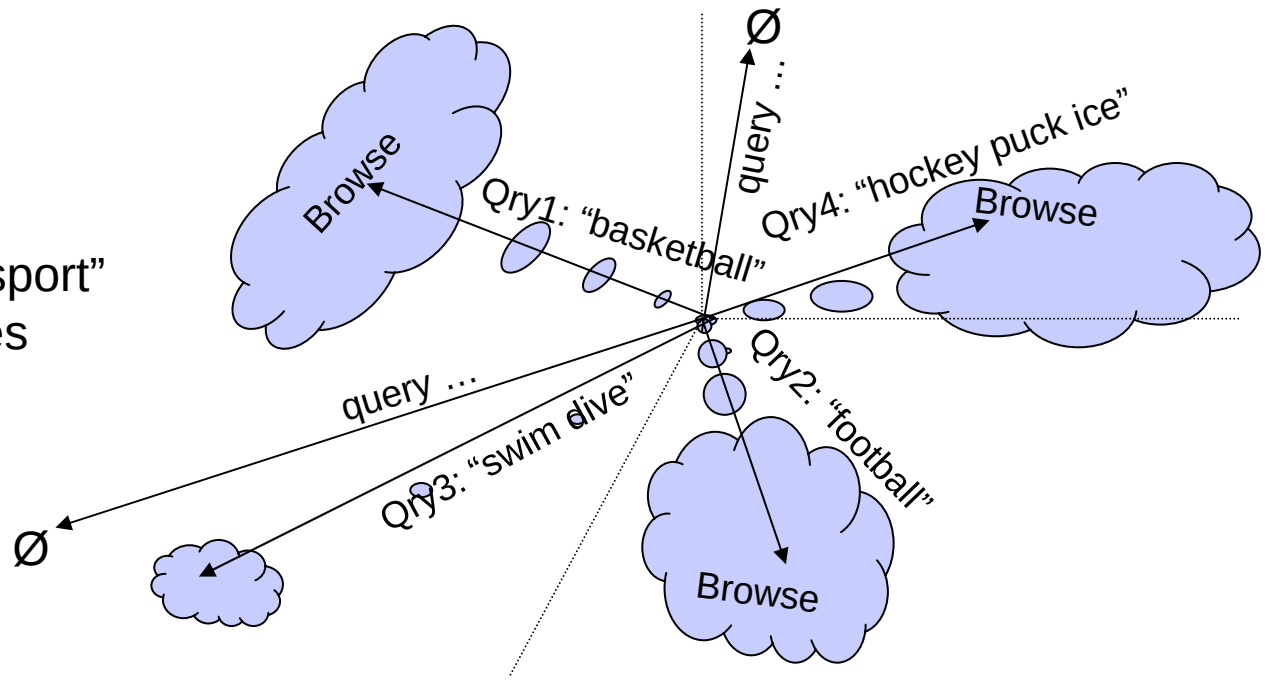
- Input: Multimodal topic, textual + image/video samples
- *Task: Acquire this Previously Unseen Semantic Concept*
- Measures system (!) performance
  - No human in the loop
  - System topic->query conversion capability
  - System learning capability
  - System search capability
- Final result is test-set independent, scalable

Samples should include **only relevant shots/images** (Dow Jones Showing Rise)

© 2004 IBM Corporation

Interactive Search = a Combination of Search and Browse

- E.g., searching for “sport” using multiple queries



A travel in video shots space

- Search: travel across the entire space, collect many shots
- Browse: local, in a neighborhood: storyboard, CBIR, etc.

© 2004 IBM Corporation

Search and Browse

	Search	**Browse**
Starting point	Query	Result list/table
Extent	Global	Local, “show similar ones”
Hits	Query matches	Anything the user can find
Collecting hits	Many, ranked	One at a time, unranked*
Processor	Computer	Human
Time	Short	Long
Contribution to MAP	?	?

© 2004 IBM Corporation

Browsing

- Browsing, “find more like this”
- Helps in query refinement
- Find and pick new matches
- Feedback, shots elevation

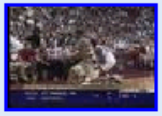
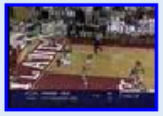
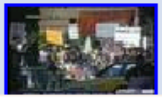
© 2004 IBM Corporation

Multimodal Query: *Concept ASR/CC phonetic VOCR CBIR* sport~ & basketball & (N.B.A.# | NBA$) & 04.38333

Results for: '**sport~ & basketball & (N.B.A.# | NBA$) & 04.38333**' in '**t04_test_wd**' Found **2178 results** Use collection shots file: **E:\HTTP\\cgi-bin\T04SearchRef.txt**. Display first 100 hits

Non-Vocabulary words: ' NBA.'. Use a # suffix for phonetic search, $ for VOCR.

Save marked results as [ ] Save | Feedback | Pseudo | Refresh

No#	Score	Video Segment	QUERY Shots
1 ○	254.79	19981213_CNNa 00:21:41-00:21:49 (basketball _&_sport)((basketball _&_CBI)_&_sport)(sport...	RefShot#: 44041 0:21:39-0:21:45 basketball RefShot#: 44042 0:21:45-0:21:47 RefShot#: 44043 0:21:47-0:21:50
2 ○	222.78	19981204_CNNa 00:21:56-00:22:03 (basketball _&_CBI) basketball (CBIR) _&_basketball)((CBI...	RefShot#: 38333 0:21:56-0:21:57 basketball  RefShot#: 38334 0:21:57-0:22:01  RefShot#: 38335 0:22:01-0:22:04
3 ○	206.92	19981002_CNNa 00:16:50-00:16:58 /_/_/M/IH/N/T/IX/DD/ _/N/B/AE/ _/B/R/UW/S/ _/F/R/(spor...	RefShot#: 997 0:16:29-0:16:54 RefShot#: 998 0:16:54-0:17:01
4 ○	180.15	19981120_CNNa 00:21:44-00:22:13 CBIRsport sport(sport _&_CBI) (CBIR_&_sport) sport sport ...	RefShot#: 29385 0:21:43-0:21:44 RefShot#: 29386 0:21:44-0:21:50 RefShot#: 29387 0:21:50-0:21:52 RefShot#: 29388 0:21:52-0:21:57 RefShot#: 29389 0:21:57-0:22:01 RefShot#: 29390 0:22:01-0:22:07 RefShot#: 29391 0:22:07-0:22:09 RefShot#: 29392 0:22:09-0:22:16
5 ○	177.65	19981026_CNNa 00:03:08-00:03:21 (CBIR_&_sport)(sport _&_CBI)	RefShot#: 11983  RefShot#: 11984 RefShot#: 11985

© 2004 IBM Corporation

The impact of Shot Elevation on MAP

Simulated scenario: For each topic,

- Take final Interactive query as baseline list
- Elevate all the *correct* shots to the top, up to depth n

Plot MAP(n)

A random result list of length K (=1000) would yield a linear expected AP(n) of

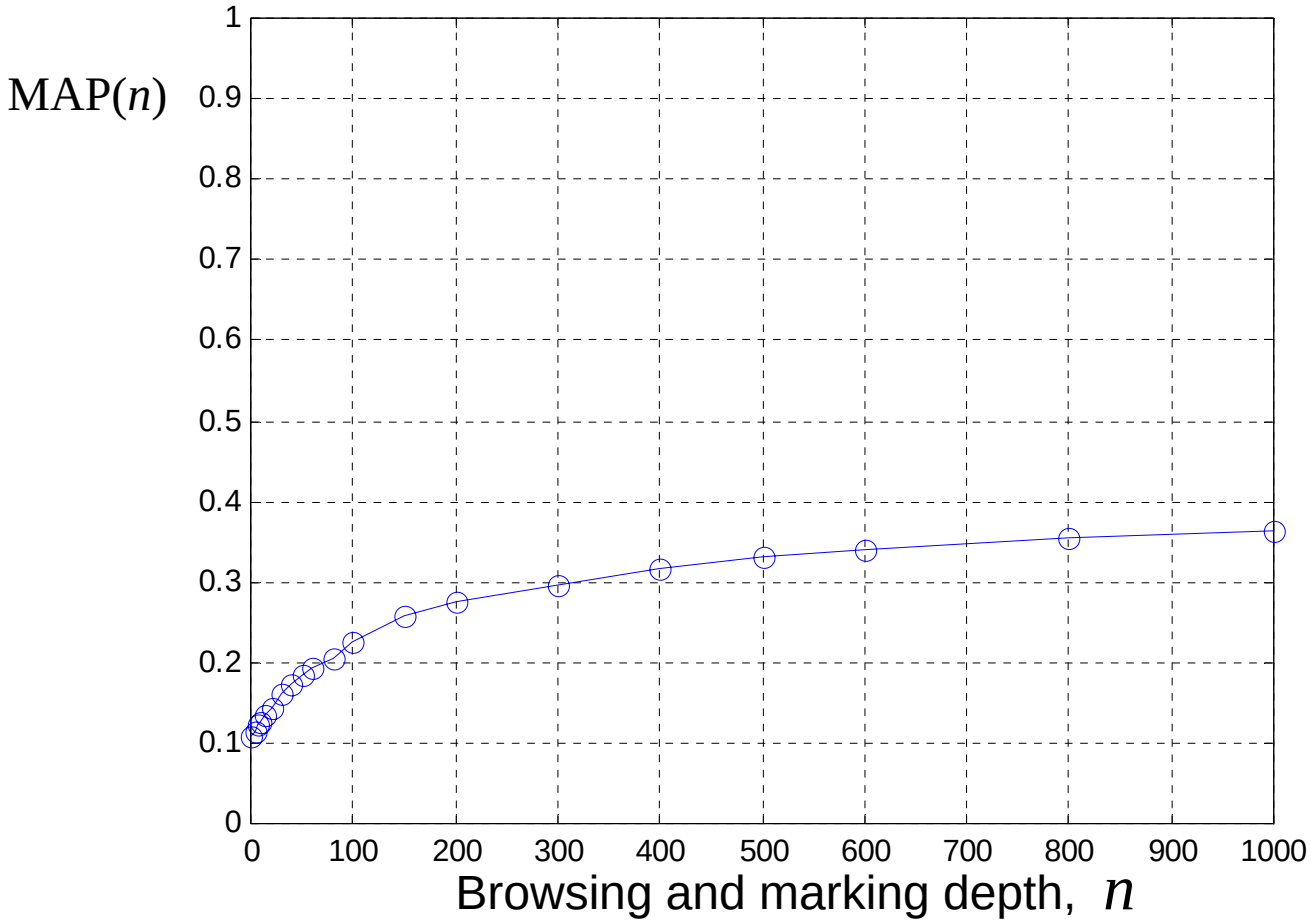
$$\overline{MAP}(n) = \frac{1}{|GT|}\left(np \cdot 1.0 + p \sum_{i=n+1}^{K} \frac{np + (i-n)p}{i} \right)$$

$$= \frac{1}{|GT|}\left(n(p - p^2) + p^2 K \right)$$

$$\text{where} \quad p = \frac{|GT|}{N}$$

© 2004 IBM Corporation

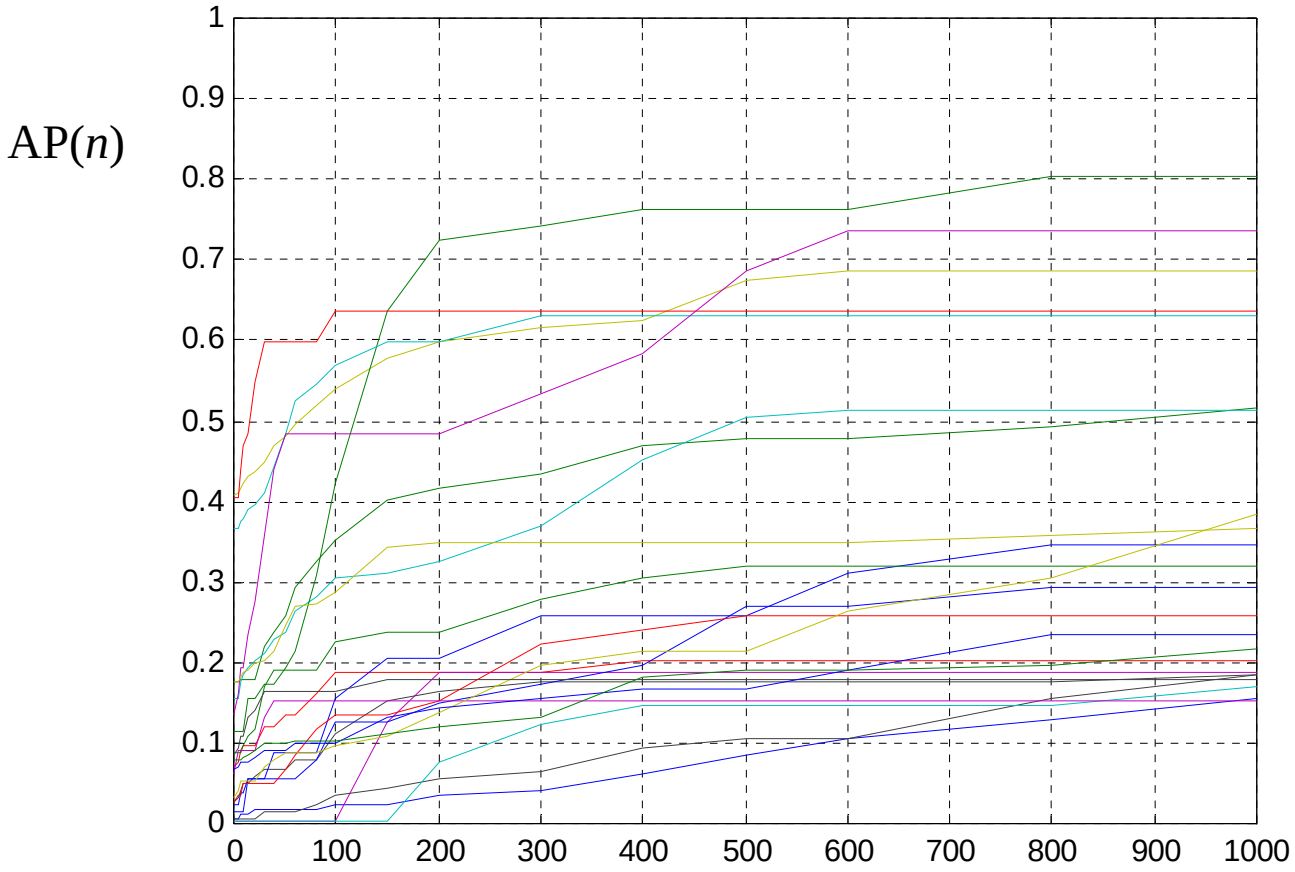
Shot Elevation Impact on MAP – Simulation Result



- Above linear -> means the search results are better than random.
- Labelling the top 100 gives more than 100% improvement in MAP

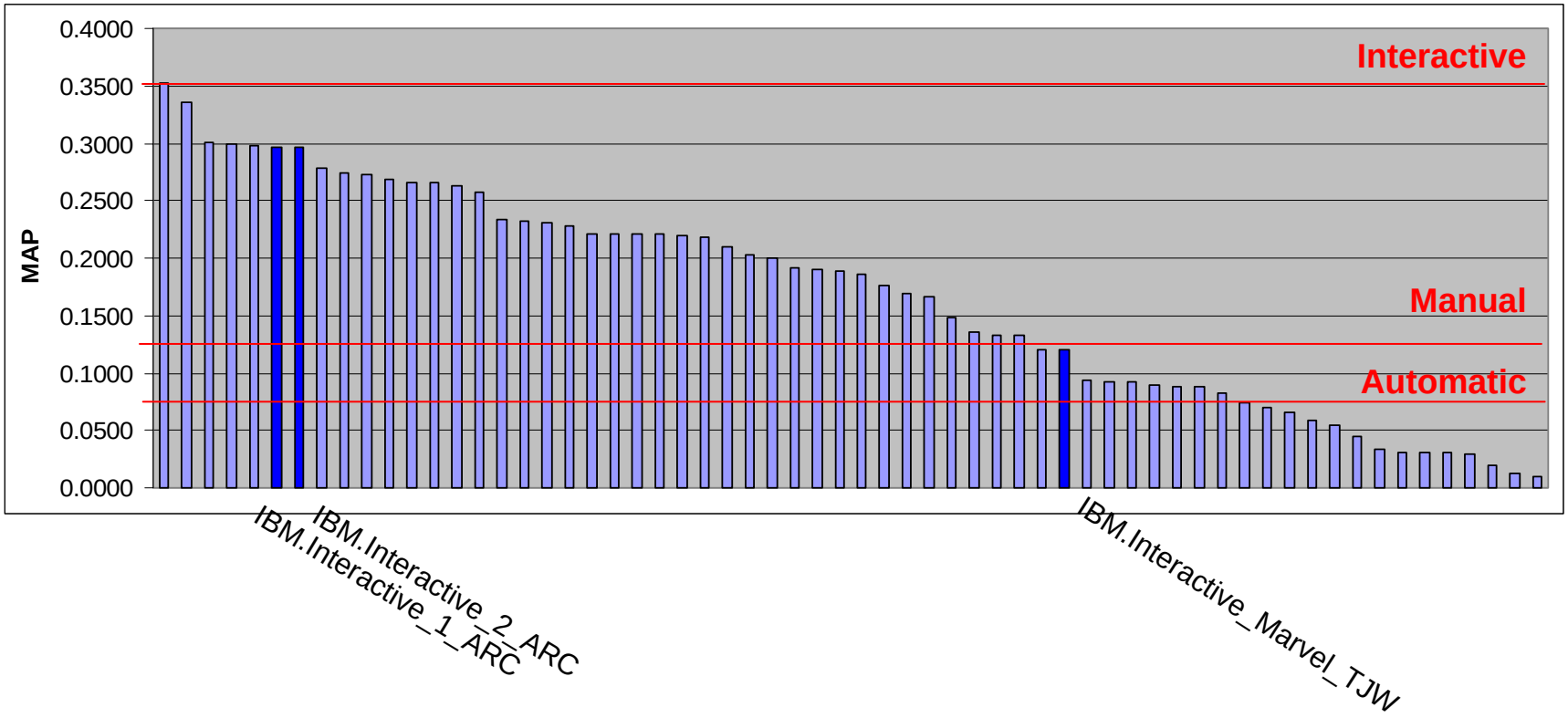
© 2004 IBM Corporation

Shots Elevation Impact on Topics AP



© 2004 IBM Corporation

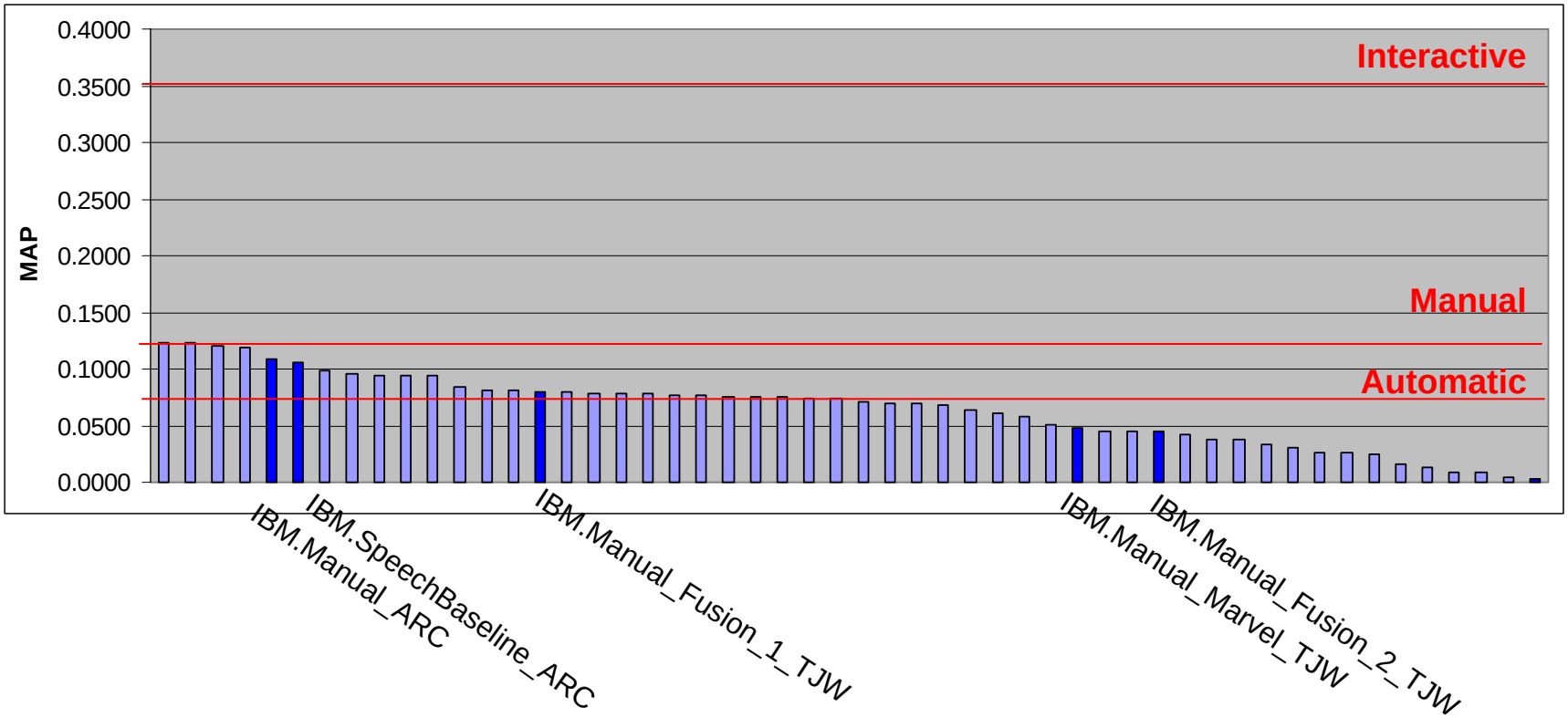
Interactive Search Runs



- Results can get much higher than manual search
- Does it scale up for large data sets ?

© 2004 IBM Corporation

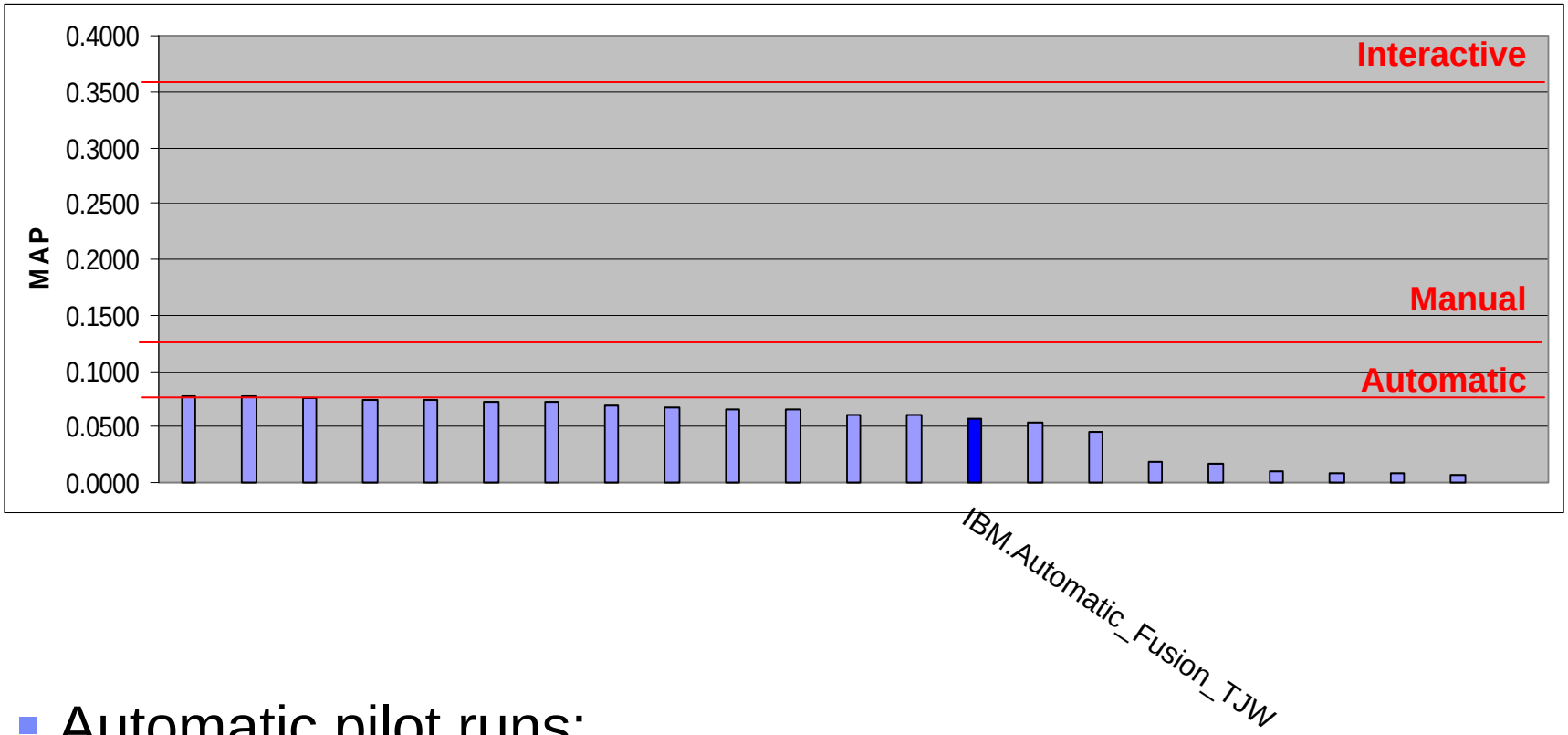
Manual Search Runs



- Scalable: same query may be run on large data sets

© 2004 IBM Corporation

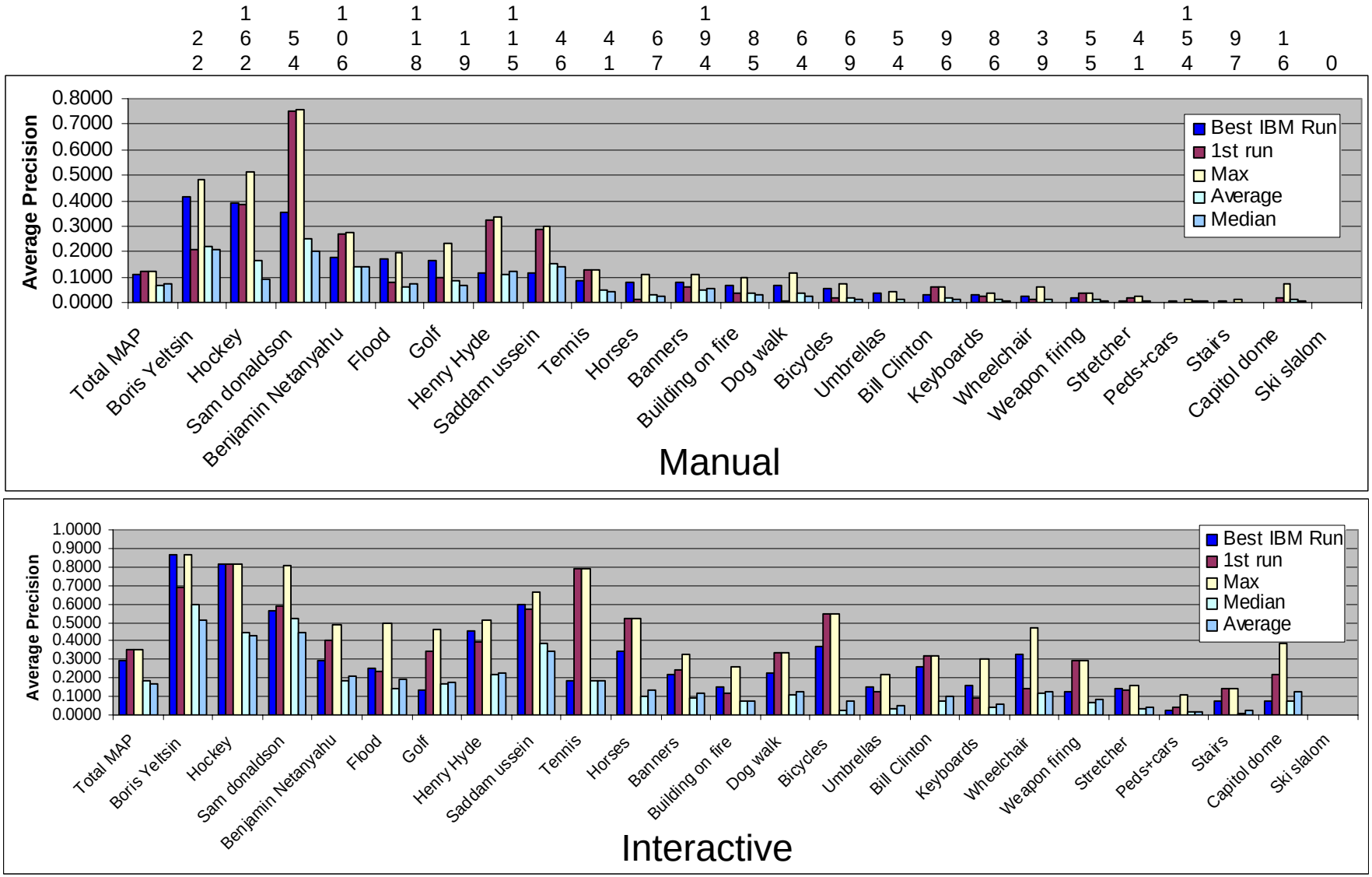
Automatic Search Runs



- Automatic pilot runs:
  - Results are comparable with the ballpark of Manual search

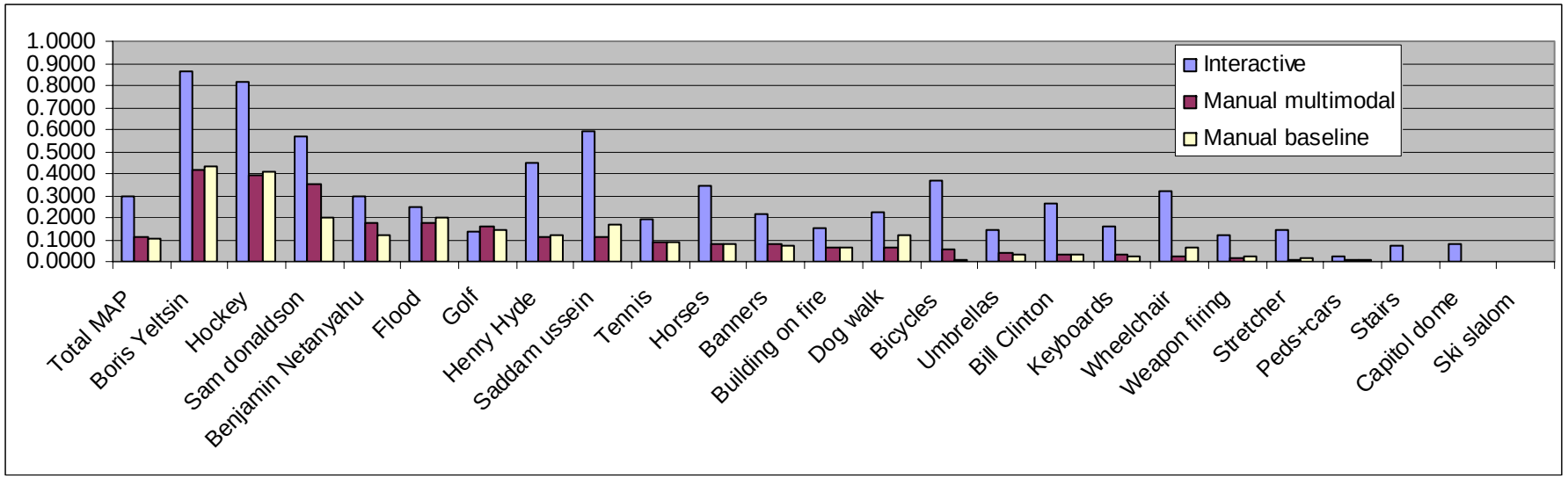
© 2004 IBM Corporation

AP by Topics



© 2004 IBM Corporation

Comparison – Interactive vs. Manual Runs



Topics are sorted by decreasing Manual AP order

© 2004 IBM Corporation

Conclusions

- Browsing and shots elevation are essential in Interactive search
- Relevance Feedback found very helpful in query formulation
- Automatic search – concept acquisition, the most objective, comparable to manual, scales up.
- Manual search – less objective, scales up, speech is still the most dominant part
- Interactive search – very interactive, subjective, CBIR is very instrumental, may not scale up as well.

© 2004 IBM Corporation