

# Passage Retrieval using HMMs

HARD 2004

University of Illinois at Urbana-Champaign

Jing Jiang

ChengXiang Zhai

## Motivation – Variable Length Passages

Nokia, the world's  
biggest ... acquired Sega  
... Japanese video game  
maker, ...  
...  
its mobile N-Gage game  
...  
... features of a cell phone,  
MP3-player ...

Nokia is the cell phone  
market leader ...

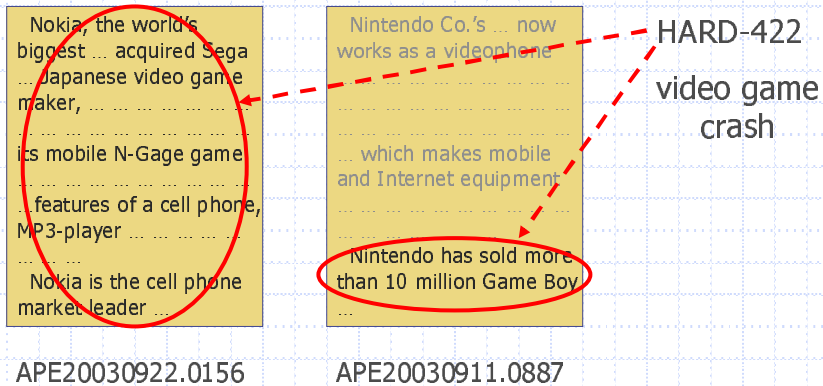
APE20030922.0156

Nintendo Co.'s ... now  
works as a videophone  
...  
... which makes mobile  
and Internet equipment  
...  
...  
Nintendo has sold more  
than 10 million Game Boy  
...

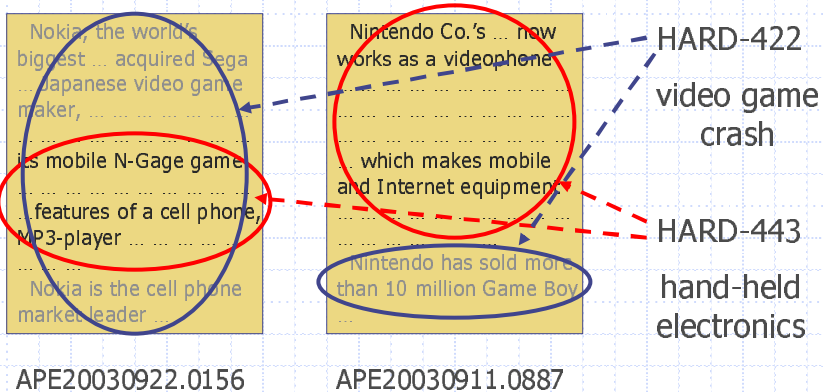
Nintendo has sold more  
than 10 million Game Boy  
...

APE20030911.0887

# Motivation – Variable Length Passages document-dependent



# Motivation – Variable Length Passages query-dependent



## Research Question

- ◆ Passage length is
  - document-dependent
  - query-dependent

How to detect  
variable-length passages?

## Previous Work on Passage Retrieval

- ◆ Structural or semantic boundary
  - Passage is not query-specific.
- ◆ Fixed-length
  - Passage length is not query-specific.
  - Passage content may not be coherent.
- ◆ Arbitrary – MultiText
  - Only query words are considered.
  - Heuristics are used to reduce search space.
- ◆ HMM-based
  - The method is promising, but previous work didn't fully explore its potential.

# HMM-Based Method

document

w | w | ... w | w | w | w | ... w | w | w | w | ... w

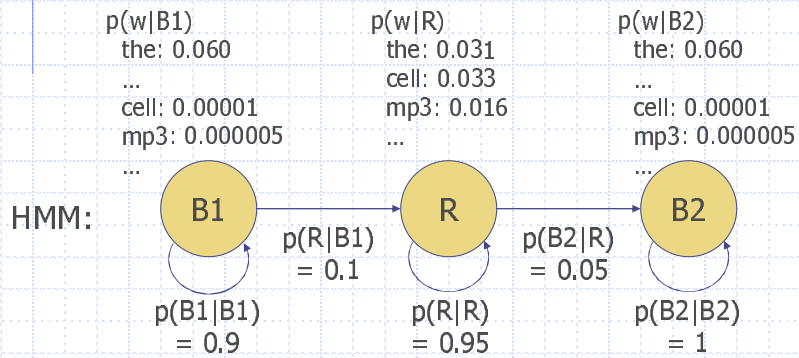
# HMM-Based Method

Q: hand-held electronics

document

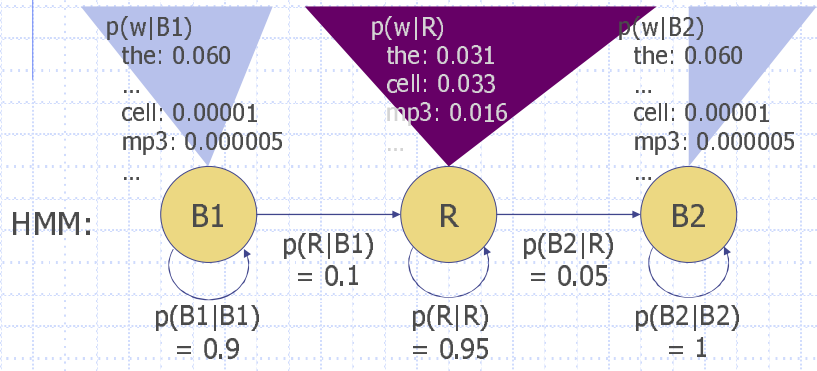
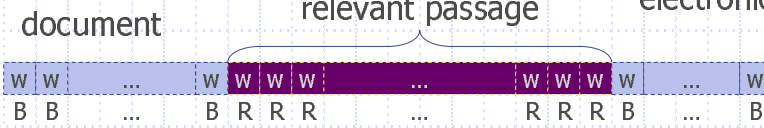
relevant passage

w | w | ... w | w | w | w | ... w | w | w | w | ... w

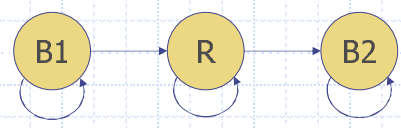


# HMM-Based Method

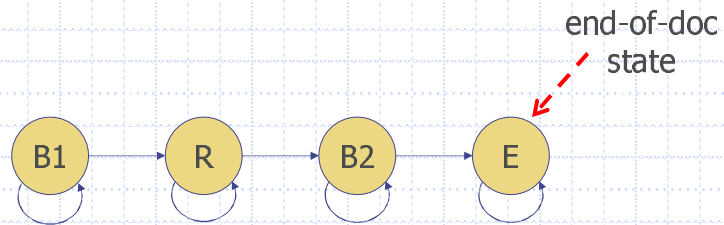
Q: hand-held electronics



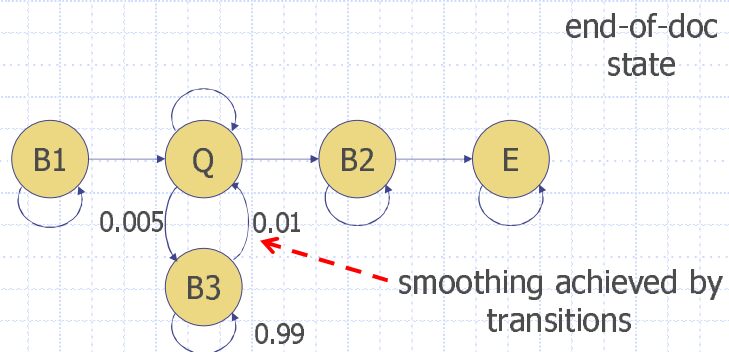
# Constructing the HMM



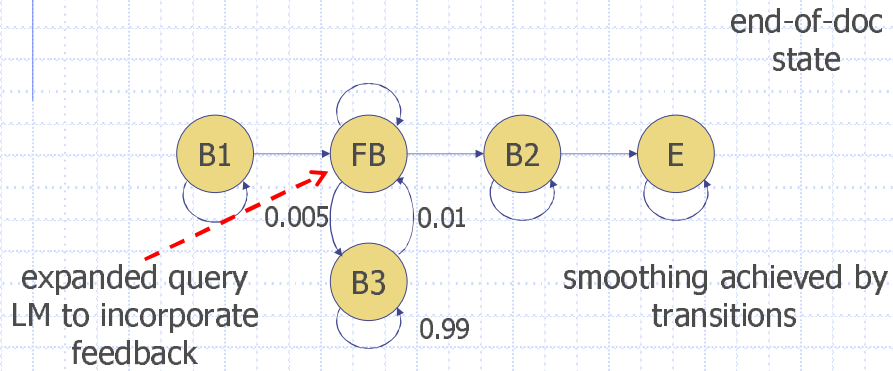
## Constructing the HMM



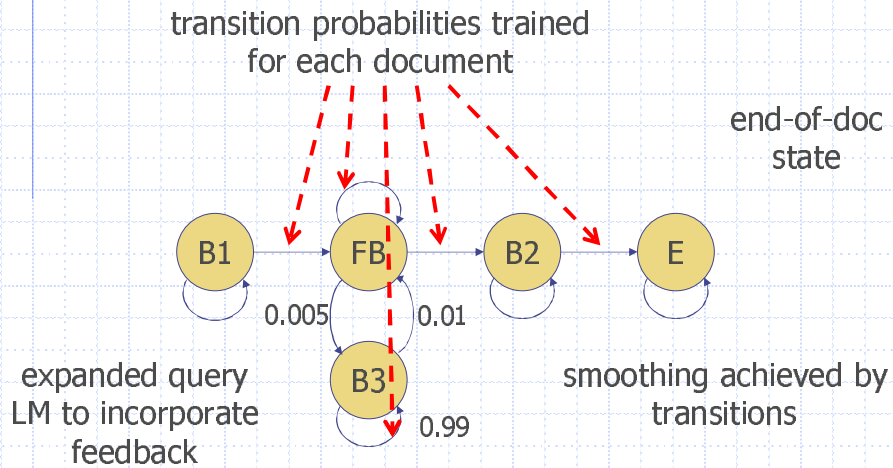
## Constructing the HMM



# Constructing the HMM

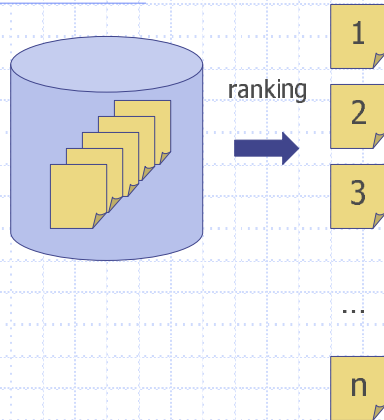


# Constructing the HMM

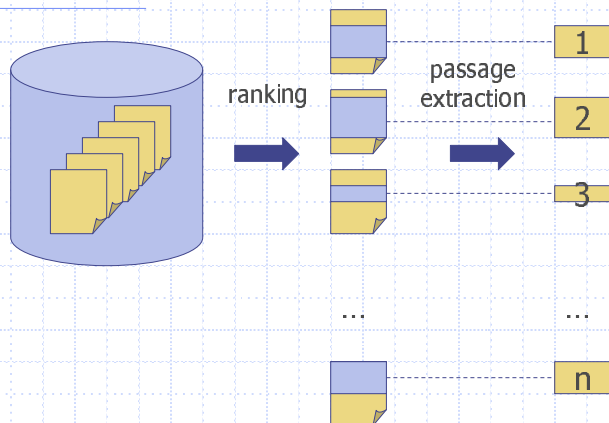




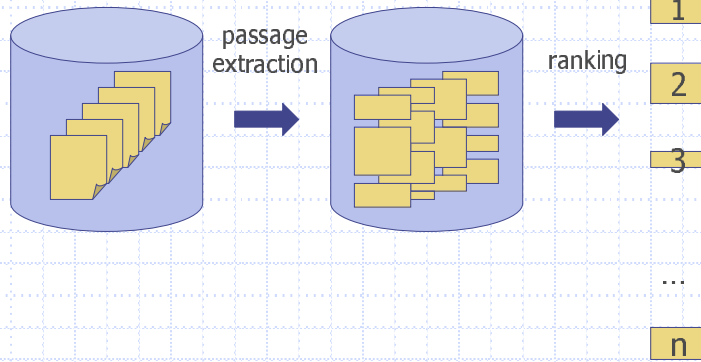
## Retrieval – Approach 1



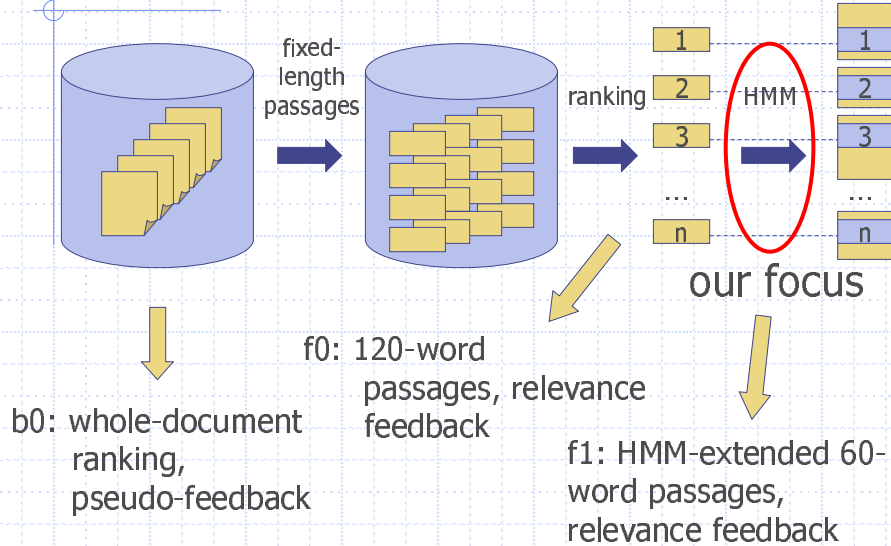
## Retrieval – Approach 1



## Retrieval – Approach 2



## Retrieval – Our Approach



## Passage-Level Results

Run	Explanation	BPref @ 12K chars	Rec @ 10 psgs	Prec @ 10 psgs
b0	whole-doc, pseudo FB	<b>0.2710</b>	<b>0.2517</b>	<b>0.1570</b>
f0	fixed 120 psgs, rel FB	0.2080	0.1067	0.2391
f1	fixed 60 psgs + HMM, rel FB	0.1860	0.1494	0.1411

◆ Overall, baseline was the best.

## Effectiveness of HMM method

Method	BPref @ 12K	Prec @ 12K	CharRPrec
Fixed 60	0.1208	0.1623	0.0776
Fixed 60 + HMM	0.1868	0.2143	0.1424
Relative improvement	54.6%	32.0%	83.5%
Fixed 120	0.1738	0.2088	0.1043
Fixed 120 + HMM	0.2131	0.2265	0.1562
Relative improvement	22.6%	8.48%	49.8%

- ◆ HMM method improved performance over fixed-length passages
- ◆ Less improvement if fixed-length closer to optimal length

## Diagnosis Runs

Factor	Feedback	Ranking	HMM	Overall
b0	pseudo FB	doc	yes	N/A
f1	rel FB	passage	no	N/A
f1 vs. b0 (BPref@12K)	-12.1%	-29.7%	+10.0%	-32.1%

non-optimal  
parameter setting

KL-divergence works  
poorly on passages

HMM improves  
boundaries

## Discussions and Conclusions

- ◆ HMM method improved the performance over fixed-length passages
- ◆ LM (KL-divergence) method gives worse performance on passage ranking than on document ranking

The End

◆ Questions?