Transform Methods for Information Retrieval

Laurence A. F. Park

lapark@ee.mu.oz.au

ARC Special Research Centre for Ultra-Broadband Information Networks

THE UNIVERSITY OF MELBOURNE

Outline

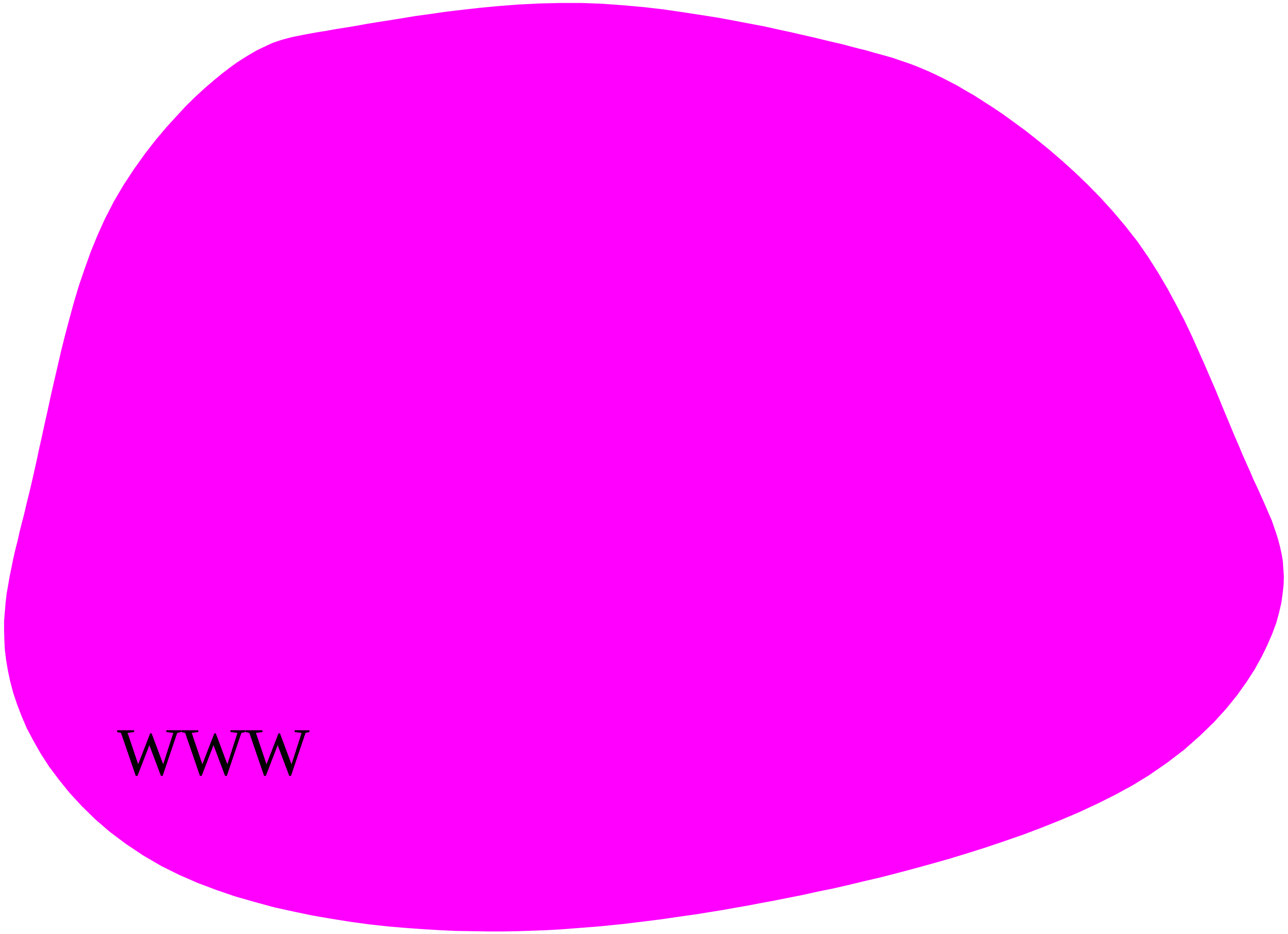
- Why do we need text retrieval?
- Our method : Spectral Document Ranking (SDR)
- Improving Efficiency of SDR
- Multiresolution Analysis with SDR
- Comparisons

Outline

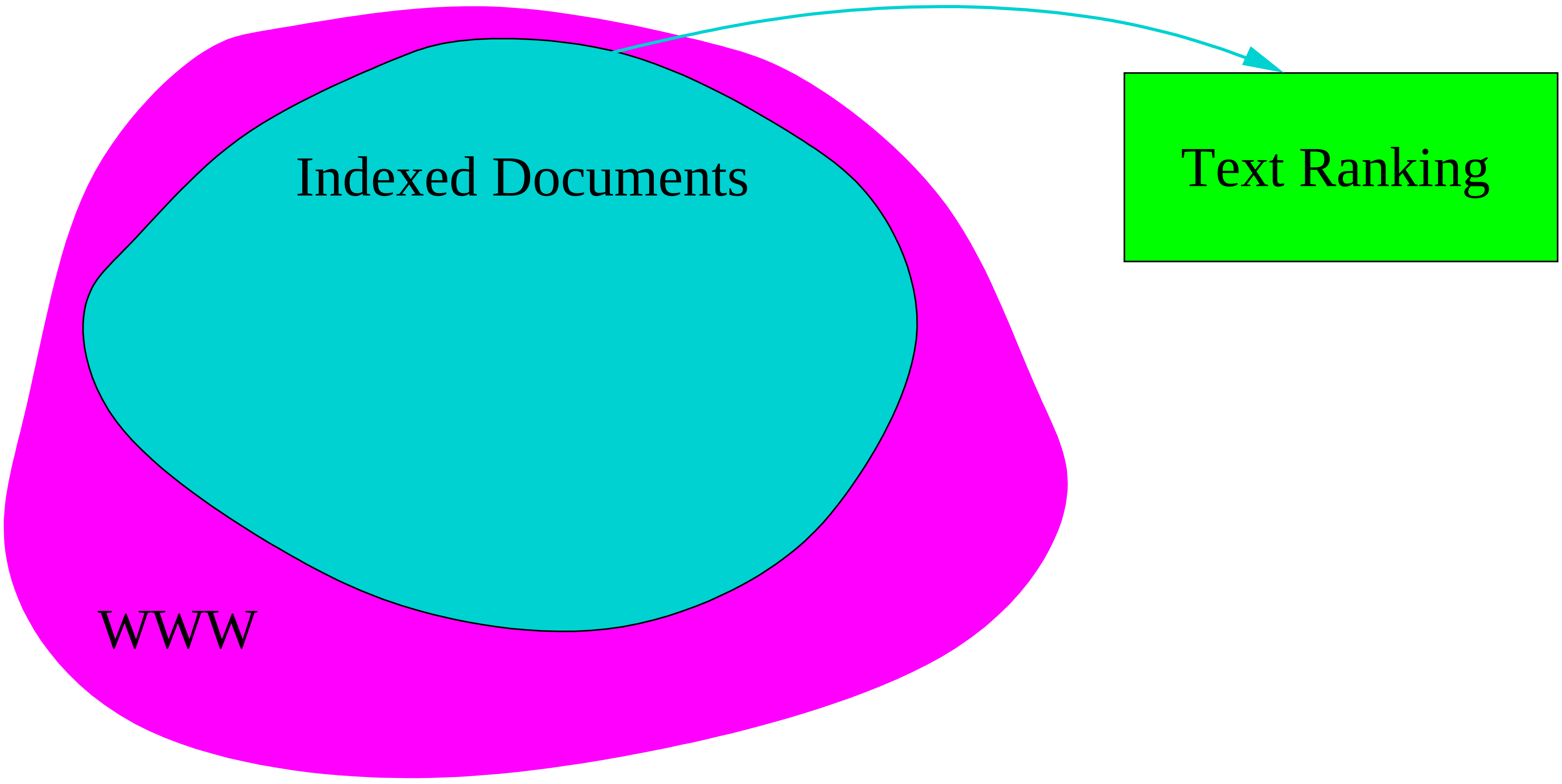
- Why do we need text retrieval?
  - WWW problem
  - Vector Space Methods
  - Proximity Methods
  - Experimental Data
- Our method : Spectral Document Ranking (SDR)
- Improving Efficiency of SDR
- Multiresolution Analysis with SDR
- Comparisons

CUBIN

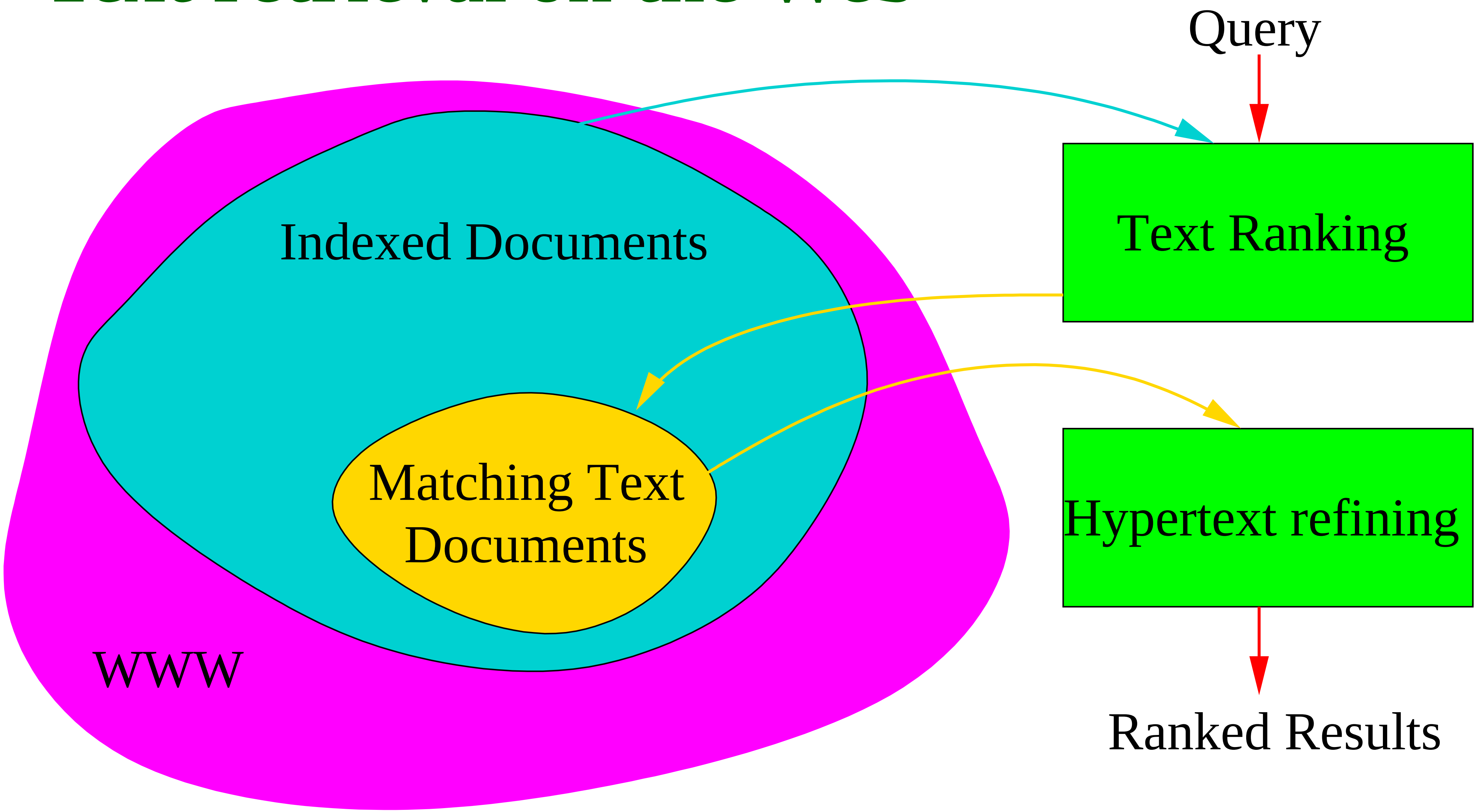
Text retrieval on the Web



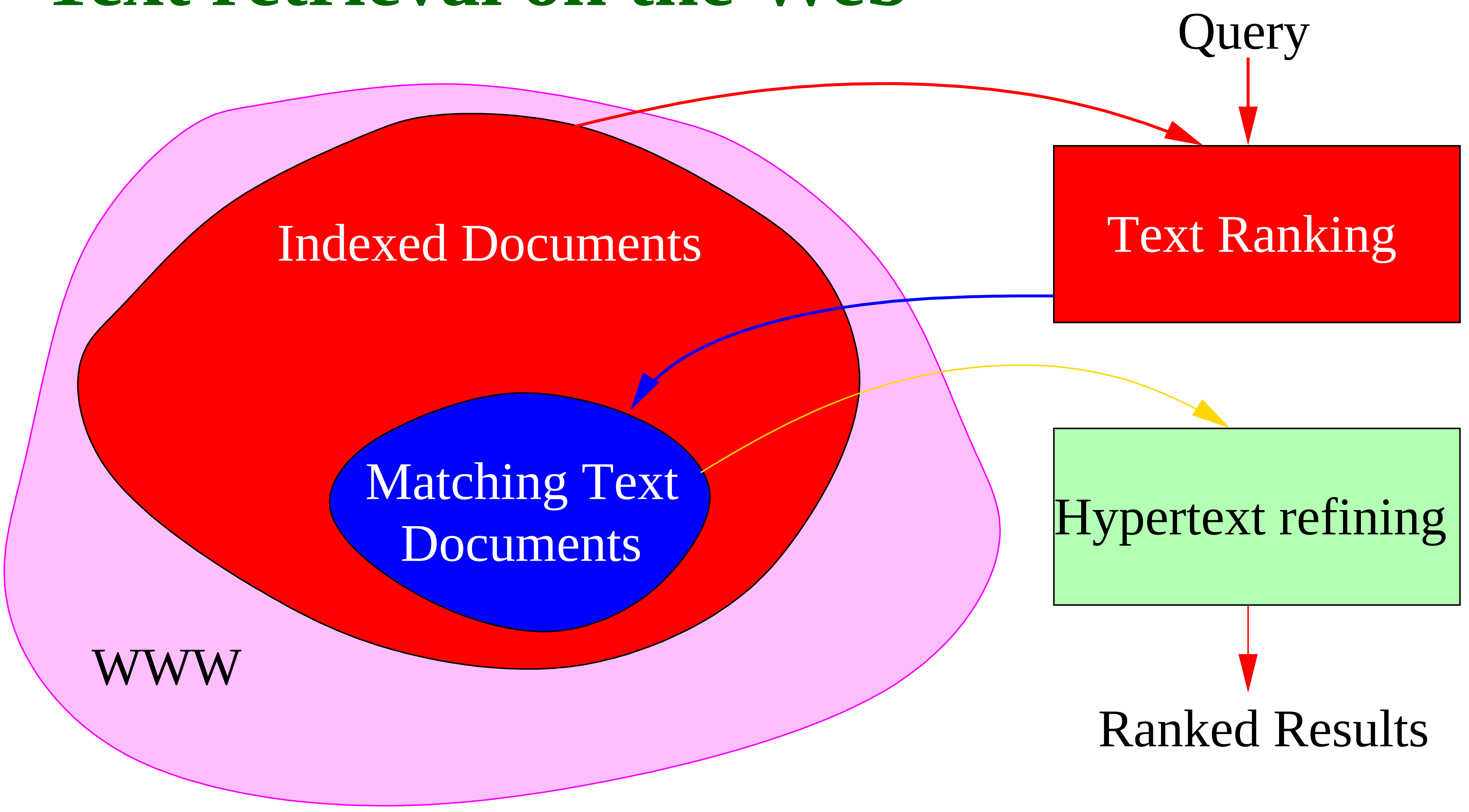
Text retrieval on the Web

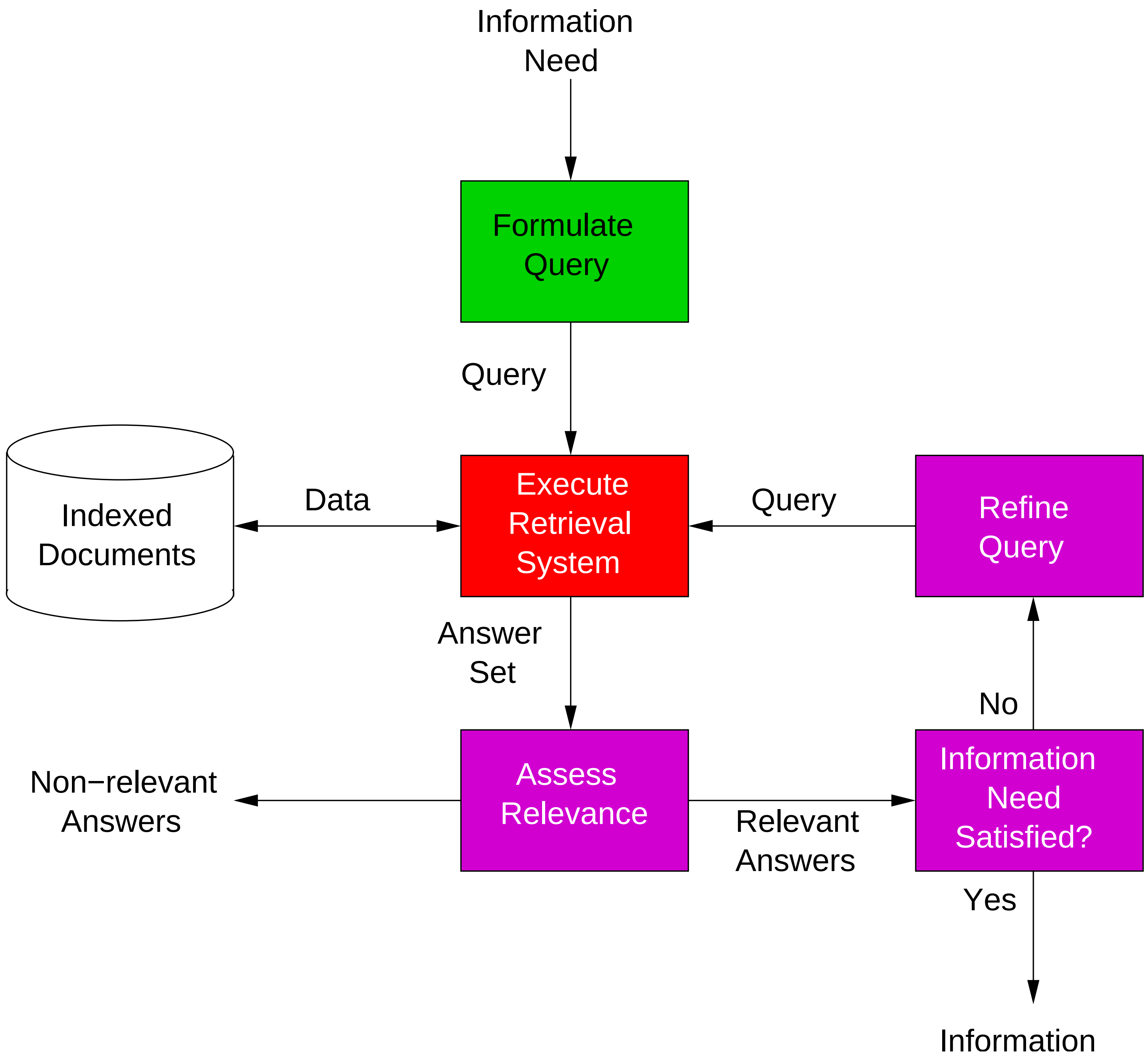


Text retrieval on the Web



Text retrieval on the Web




Information
Need
Formulate
Query
Query
Indexed
Documents
Data
Execute
Retrieval
System
Query
Refine
Query
Answer
Set
No
Non−relevant
Answers
Assess
Relevance
Relevant
Answers
Information
Need
Satisfied?
Yes
Information

Issue of Relevance

What makes a document relevant to a set of query terms?

- The occurrence of the terms in the document?
- The proximity of the terms?

Issue of Relevance

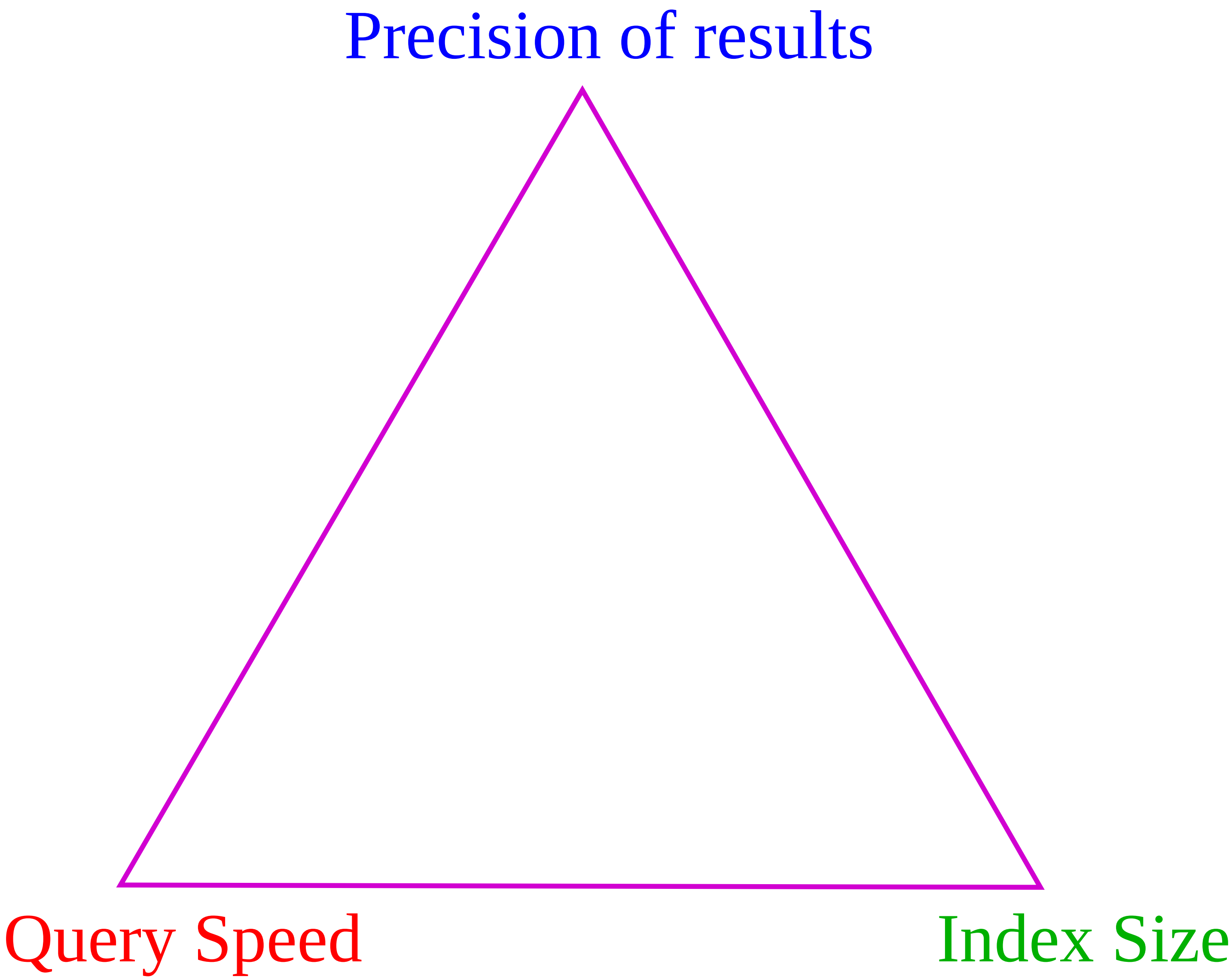
What makes a document relevant to a set of query terms?

- The occurrence of the terms in the document?
- The proximity of the terms?

Relevance is subjective, therefore we must:

- Give documents a score of relevance
- Present the ordered list to the searcher

Important Properties



Existing method: Vector Space Method

	term 1	term 2	…	term m
document 1	0	2	…	0
document 2	1	1	…	0
⋮	⋮	⋮	⋱	⋮
document n	0	1	…	3
query	0	1	…	1

- Document Score = query vector • document vector
- ✓ Fast and compact
- ✗ Ignores query term proximity within documents

Existing method: Proximity Method

Eg. Smallest Substring:

Happy^{1} birthday^{2} to^{3} (you^{4},
happy^{5} birthday^{6}) to^{7} (you^{8},
happy^{9} birthday^{10}) Mr.^{11} President^{12},
(happy^{13} birthday^{14} to^{15} you^{16})

Query: "happy you birthday" = $\{(4,6),(8,10),(13,16)\}$

- Document score = function(query term position or proximity)
- ✓ Higher quality results
- ✗ Slow, more storage needed

Experimental Data

- TREC (Text REtrieval Conference) at NIST
- Text articles from:
  - Associated Press 1988 (79,919 documents)
  - Wall Street Journal 1990 (74,520 documents)
  - Computer Select (56,920 documents)
  - Federal Register 1988 (19,860 documents)
- Queries from:
  - TREC-1 Conference (50 queries)
  - TREC-2 Conference (50 queries)
  - TREC-3 Conference (50 queries)
- Selected titles to simulate typical Web query

Outline

- Why do we need text retrieval?
- Our method : Spectral Document Ranking (SDR)
  - Term signals
  - Fourier Domain Scoring
- Improving Efficiency of SDR
- Multiresolution Analysis with SDR
- Comparisons

Spectral Document Ranking

- Vector space methods ignore spatial data
- Proximity methods are slow (many comparisons)
- Rank documents according to query term spectra:
  - Spectral components are orthogonal
  - Magnitude portrays query term occurrence
  - Phase shows component position
- Query term spectra calculated from query term signals

Creating a term signal : huntsman

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$$\tilde{f}_{d,huntsman} = \begin{array}{|c|c|c|c|c|c|c|c|}\hline \quad & \quad & \quad & \quad & \quad & \quad & \quad & \quad \\ \hline \end{array}$$

Creating a term signal : huntsman

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$$\tilde{f}_{d,huntsman} = \boxed{}\boxed{}\boxed{}\boxed{}\boxed{}\boxed{}\boxed{}\boxed{}$$

Creating a term signal : huntsman

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$$\tilde{f}_{d,huntsman} = \begin{array}{|c|c|c|c|c|c|c|c|} \hline 1 & & & & & & & \\ \hline \end{array}$$

Creating a term signal : huntsman

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$$\tilde{f}_{d,huntsman} = \begin{array}{|c|c|c|c|c|c|c|c|} \hline 1 & 0 & & & & & & \\ \hline \end{array}$$

Creating a term signal : huntsman

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$$\tilde{f}_{d,huntsman} =$$

1	0	1					

Creating a term signal : huntsman

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$\tilde{f}_{d,huntsman} =$

1	0	1	0				

CUBIN

Creating a term signal : huntsman

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$\tilde{f}_{d,huntsman} =$

1	0	1	0	1			

Creating a term signal : huntsman

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$$\tilde{f}_{d,huntsman} = \begin{array}{|c|c|c|c|c|c|c|c|} \hline 1 & 0 & 1 & 0 & 1 & 0 & & \\ \hline \end{array}$$

CUBIN

Creating a term signal : huntsman

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$\tilde{f}_{d,huntsman} =$

1	0	1	0	1	0	1	

CUBIN

Creating a term signal : huntsman

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$$\tilde{f}_{d,huntsman} = \begin{array}{|c|c|c|c|c|c|c|c|}\hline 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ \hline\end{array}$$

Creating a term signal : rock

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$\tilde{f}_{d,rock} =$

Creating a term signal : rock

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$$\tilde{f}_{d,rock} = \begin{array}{|c|c|c|c|c|c|c|c|} \hline \; & \; & \; & \; & \; & \; & \; & \; \\ \hline \end{array}$$

Creating a term signal : rock

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$\tilde{f}_{d,rock} =$

0								

Creating a term signal : rock

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$$\tilde{f}_{d,rock} =$$

0	2						

Creating a term signal : rock

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$$\tilde{f}_{d,rock} = \begin{array}{|c|c|c|c|c|c|c|c|} \hline 0 & 2 & 0 & & & & & \\ \hline \end{array}$$

Creating a term signal : rock

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$\tilde{f}_{d,rock} =$

0	2	0	0				

Creating a term signal : rock

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$\tilde{f}_{d,rock} =$

0	2	0	0	1			

Creating a term signal : rock

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$\tilde{f}_{d,rock} =$

0	2	0	0	1	0		

Creating a term signal : rock

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$\tilde{f}_{d,rock} =$

0	2	0	0	1	0	0	

Creating a term signal : rock

Huntsman Spiders are usually found living under loose bark on trees, in crevices on rock walls and in logs, under rocks and slabs of bark on the ground, and on foliage. Dozens of the social huntsman species, Delena cancerides , can be seen sitting together under bark on dead trees and stumps (notably wattles) but they can also be found on the ground under rocks and bark slabs. Badge Huntsman Spiders are often found on foliage but some woodland species are burrow builders, with and without trapdoors. Huntsman spiders of many species sometimes enter houses. They are also notorious for entering cars, and being found hiding behind sun visors or running across the dashboard.

$\tilde{f}_{d,rock} =$

0	2	0	0	1	0	0	0

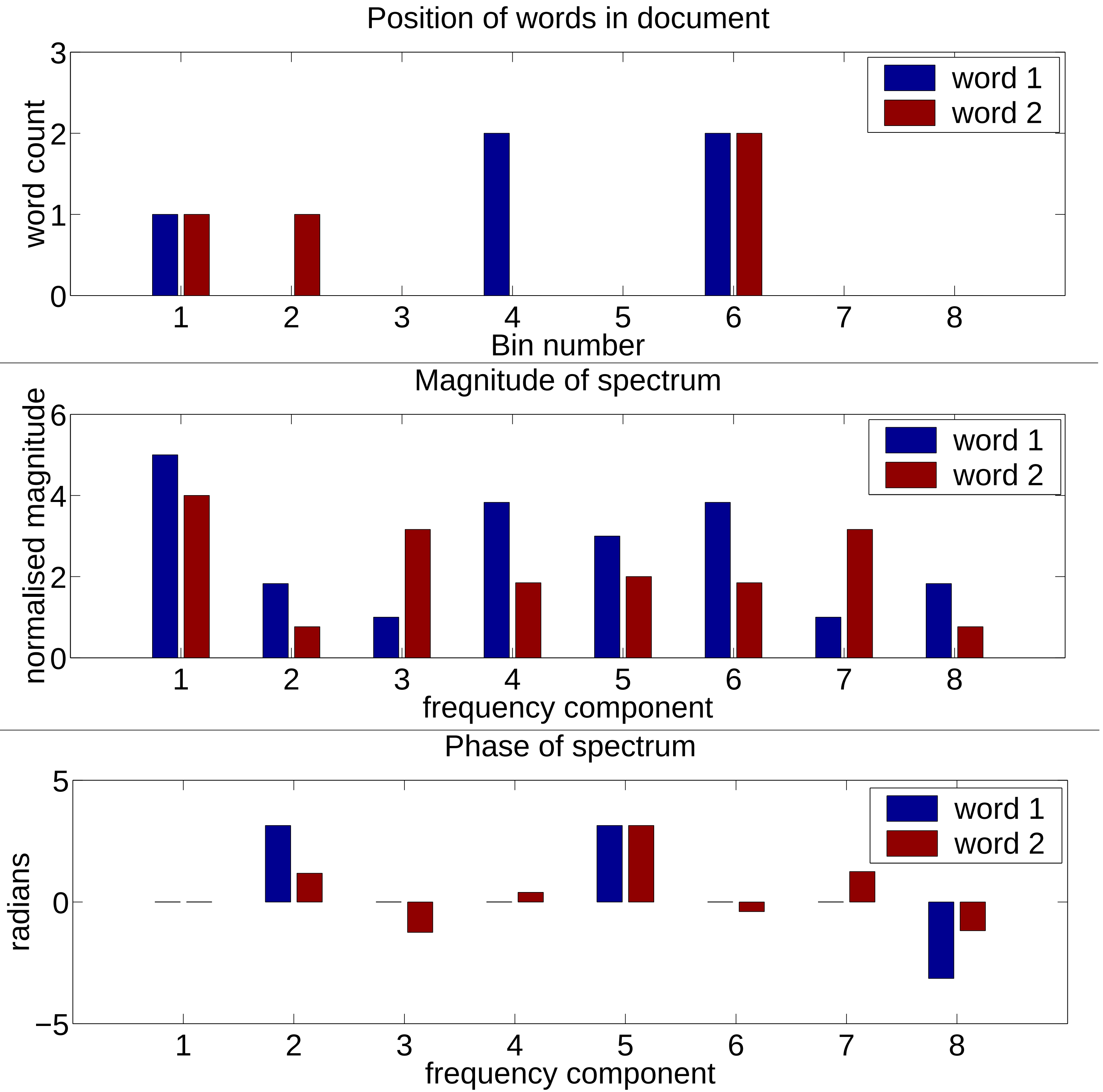
Spectral Ranking Method

After the Fourier transform:

- Term signals are decomposed into sinusoids (frequency domain)
- If term signal spectra are similar, terms appear together
- Magnitude $\propto$ occurrence of signal
- Phase = position of signal

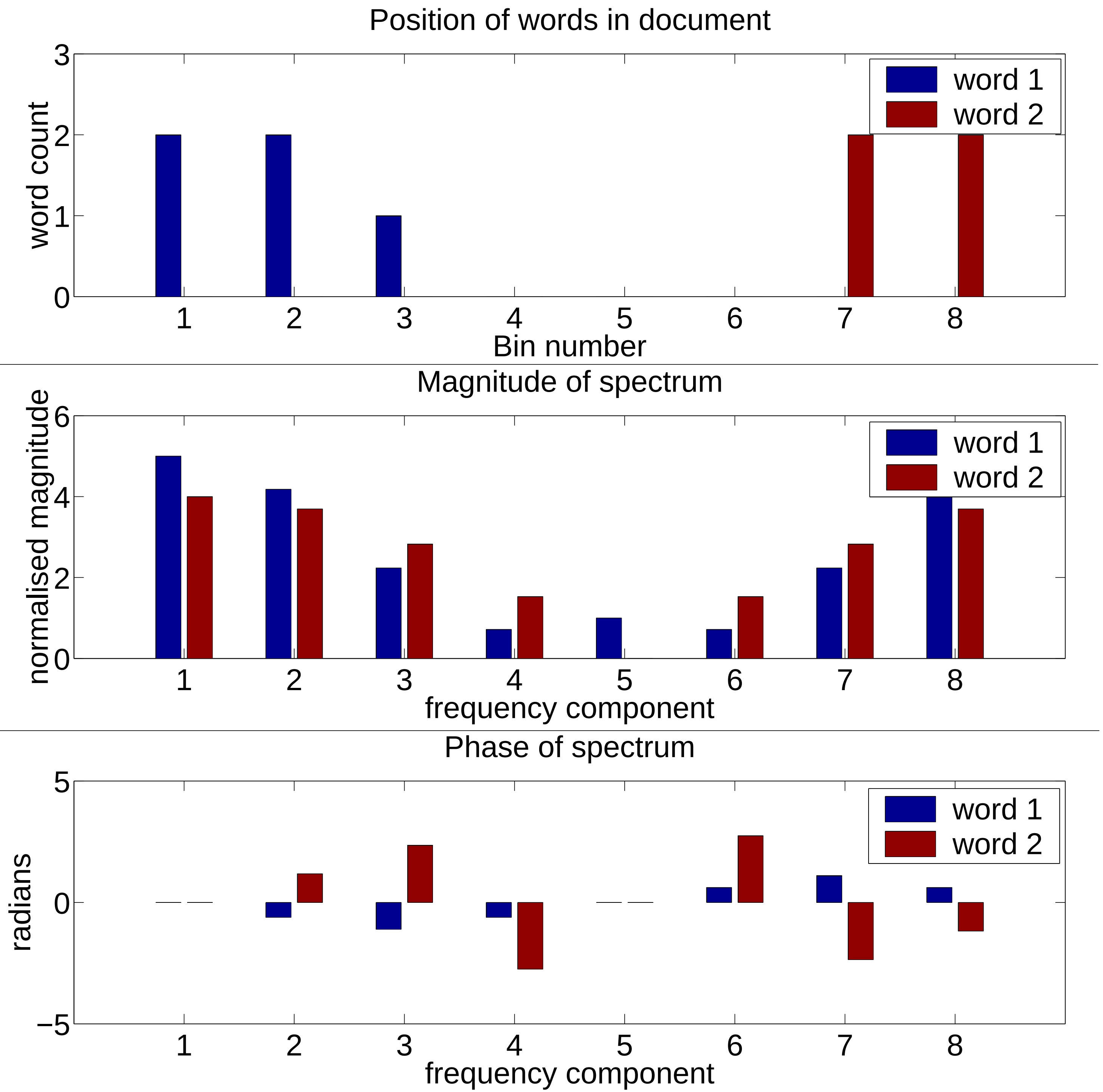
Relevant document has high magnitude across query term signals and common phase across query term signals

Close



Position of words in document
word count
Bin number
word 1
word 2
Magnitude of spectrum
normalised magnitude
frequency component
word 1
word 2
Phase of spectrum
radians
frequency component
word 1
word 2

Far



Position of words in document
word count
Bin number
word 1
word 2
Magnitude of spectrum
normalised magnitude
frequency component
Phase of spectrum
radians
frequency component

Calculate Fourier Transform

- To begin calculating the document score, we must obtain the Fourier transform of each term signal

$$\mathrm{DFT}\left(\tilde{f}_{d,huntsman}\right) = \begin{array}{|c|c|c|c|c|c|c|c|} \hline 4 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline \end{array}$$

$$\mathrm{DFT}\left(\tilde{f}_{d,rock}\right) = \begin{array}{|c|c|c|c|c|c|c|c|} \hline 3 & 1.4 & 2.2 & 2.7 & 1 & 2.7 & 2.2 & 1.4 \\ 0 & -1.2 & -1.1 & -2.6 & 3.1 & 2.6 & 1.1 & 1.2 \\ \hline \end{array}$$

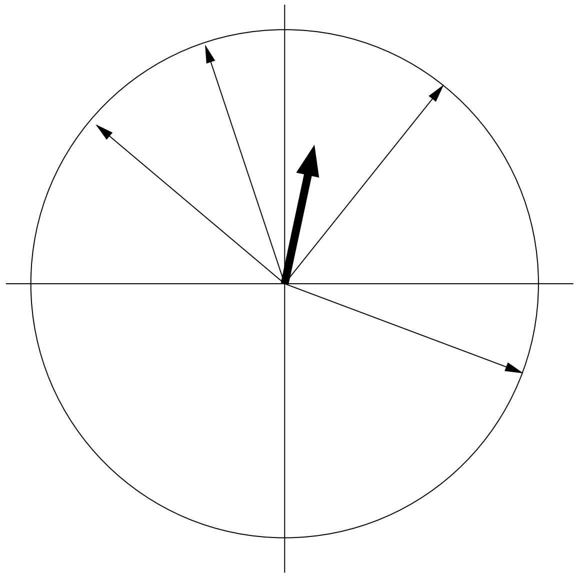
Sum Magnitudes

- The magnitude represents the occurrence of the term for a frequency
- Add the magnitudes of each component across the query terms

$\tilde{H}_{d,huntsman}$	4	0	0	0	4	0	0	0
$\tilde{H}_{d,rock}$	3	1.4	2.2	2.7	1	2.7	2.2	1.4
$\tilde{H}_{d}$	7	1.4	2.2	2.7	5	2.7	2.2	1.4

Phase Precision

Treat phase as a unit vector and find the average for each component



$\tilde{\phi}_{d,huntsman}$	0	0	0	0	0	0	0	0
$\tilde{\phi}_{d,rock}$	0	−1.2	−1.1	−2.6	3.1	2.6	1.1	1.2
$\tilde{\Phi}_{d}$	1	0.8	0.8	0.2	0	0.2	0.8	0.8

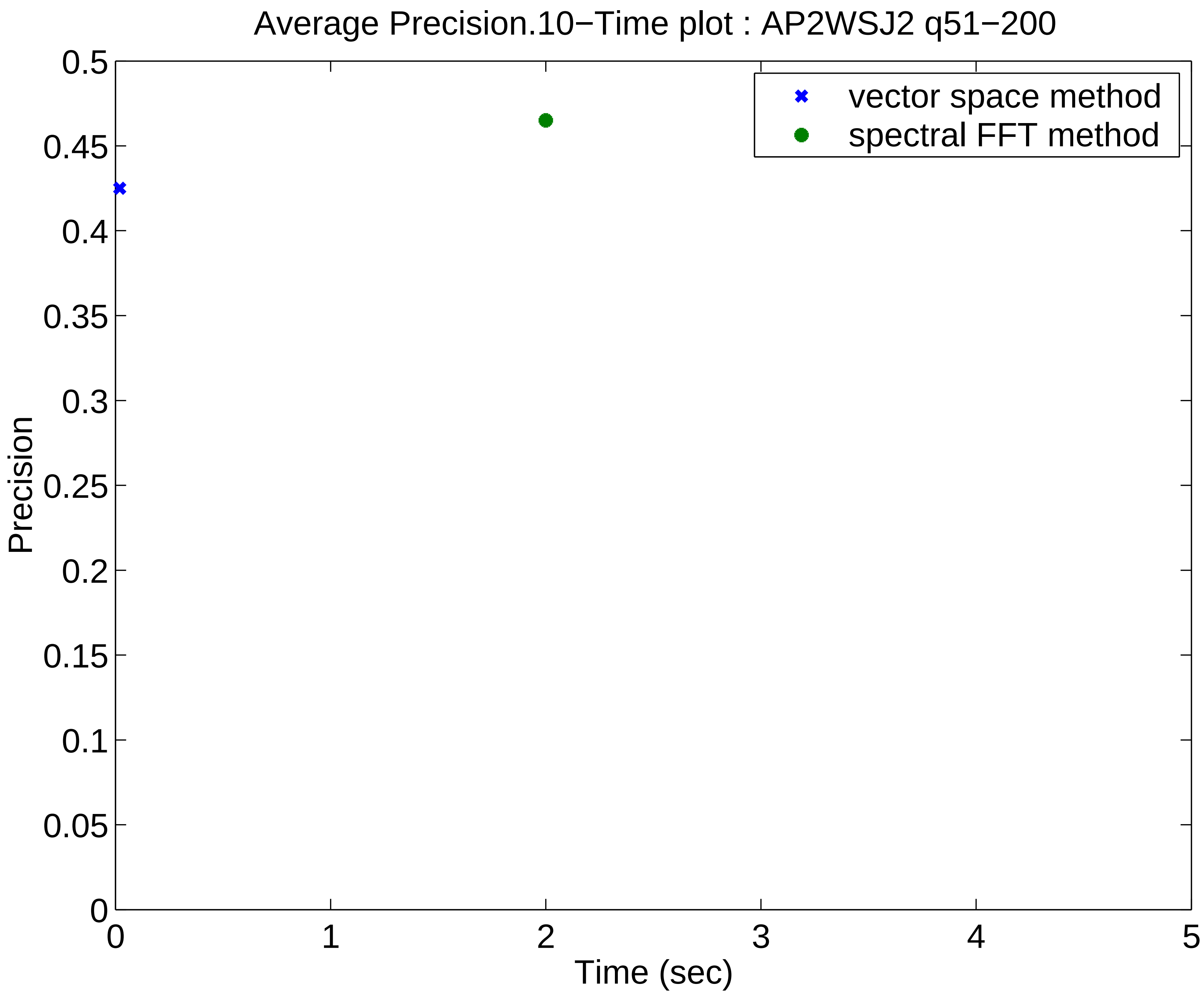
Document Score

- Multiply each magnitude by its phase precision
- Sum components to obtain document score

$\tilde{H}_d$	7	1.4	2.2	2.7	5	2.7	2.2	1.4
$\tilde{\Phi}_d$	1	0.8	0.8	0.2	0	0.2	0.8	0.8
$\tilde{s}_d$	7	1.1	1.9	0.7	0	0.7	1.9	1.1

$$\Rightarrow S_d = \sum_{b=0}^{B-1} s_{d,b} = 14.6$$

- This process is performed for each document in the set
- The documents with the highest scores are considered the most relevant by the system.


Average Precision.10−Time plot : AP2WSJ2 q51−200
Precision
Time (sec)
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
1
2
3
4
5
vector space method
spectral FFT method

CUBIN

Outline

- Why do we need text retrieval?
- Our method : Spectral Document Ranking (SDR)
- Improving Efficiency of SDR
  - Using the Cosine Transform
  - Component Cropping
  - Quantisation
  - Early Termination
- Multiresolution Analysis with SDR
- Comparisons

Problem

- We have seen the Fourier transform method provides high precision results at a slow rate
- To be competitive, we need a faster method

Problem

- We have seen the Fourier transform method provides high precision results at a slow rate
- To be competitive, we need a faster method
- Solution : Pre-compute and store the Transformed values

Problem

- We have seen the Fourier transform method provides high precision results at a slow rate
- To be competitive, we need a faster method
- Solution : Pre-compute and store the Transformed values
- Problem : DFT produces many complex floating point values

[0 0 0 1 0 0 0 0] $\overset{FFT}{\Rightarrow}$

[1 -0.7 + 0.7i -1i 0.7 + 0.7i -1 0.7 - 0.7i 1i -0.7 - 0.7i]

- Index size will increase by magnitude of 100

Problem

- We have seen the Fourier transform method provides high precision results at a slow rate
- To be competitive, we need a faster method
- Solution : Pre-compute and store the Transformed values
- Problem : DFT produces many complex floating point values

$$[\,0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0\,] \overset{FFT}{\Rightarrow}$$

$$[\,1 \quad -0.7 + 0.7i \quad -1i \quad 0.7 + 0.7i \quad -1 \quad 0.7 - 0.7i \quad 1i \quad -0.7 - 0.7i\,]$$

- Index size will increase by magnitude of 100
- Solution : Use the compression qualities of the Discrete Cosine Transform

Cosine Transform

- Advantages:
  - Decomposes signals into sinusoids
  - Karhunen-Loève Transform $\Rightarrow$ Principle Component Analysis
  - Cosine Transform = KLT for first order stationary Markov process
  - Reduction of dimension (cropping) $\Rightarrow$ Less data
  - No imaginary components
- Disadvantages:
  - Slight reduction in precision due to cropping

Compression : Component Cropping

- Cosine transform approximates KLT
- Cropping higher components removes little information

$$
\begin{aligned}
\mathrm{DCT}(f_{d,1}) &= [\,1.1 \quad 0 \quad 0.6 \quad 0.9 \quad 1.3 \quad 2.1 \quad 0 \quad 0.4\,] \\
\mathrm{DCT}(f_{d,2}) &= [\,0.8 \quad 0.3 \quad 0.7 \quad 0.7 \quad 1.1 \quad 1.0 \quad 1.5 \quad 0.9\,] \\
\mathrm{DCT}(f_{d,3}) &= [\,1.4 \quad 2.1 \quad 0.8 \quad 0.9 \quad 1.1 \quad 2.0 \quad 1.8 \quad 0.4\,] \\
\vdots \quad & \quad [\,\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots\,] \\
\mathrm{DCT}(f_{d,m}) &= [\,2.5 \quad 1.9 \quad 1.3 \quad 1.6 \quad 0.1 \quad 1.1 \quad 1.2 \quad 2.1\,]
\end{aligned}
$$

Compression : Component Cropping

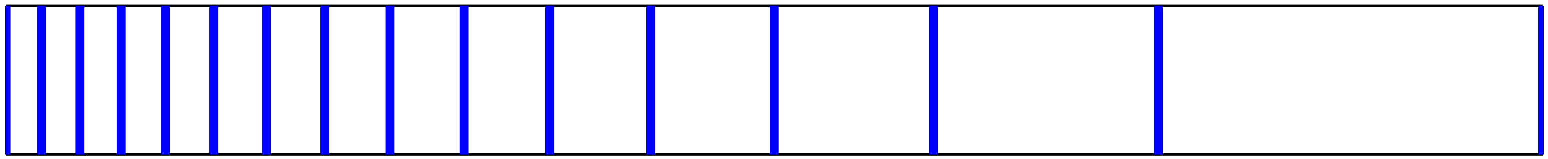
- Cosine transform approximates KLT
- Cropping higher components removes little information

$\mathrm{DCT}(f_{d,1}) =$	[1.1	0	0.6	0.9	1.3	2.1	0	0.4]
$\mathrm{DCT}(f_{d,2}) =$	[0.8	0.3	0.7	0.7	1.1	1.0	1.5	0.9]
$\mathrm{DCT}(f_{d,3}) =$	[1.4	2.1	0.8	0.9	1.1	2.0	1.8	0.4]
$\vdots$	[$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$]
$\mathrm{DCT}(f_{d,m}) =$	[2.5	1.9	1.3	1.6	0.1	1.1	1.2	2.1]

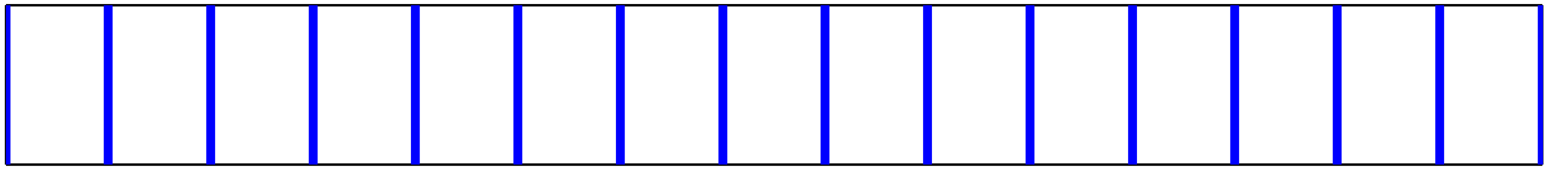
Compression : Quantisation

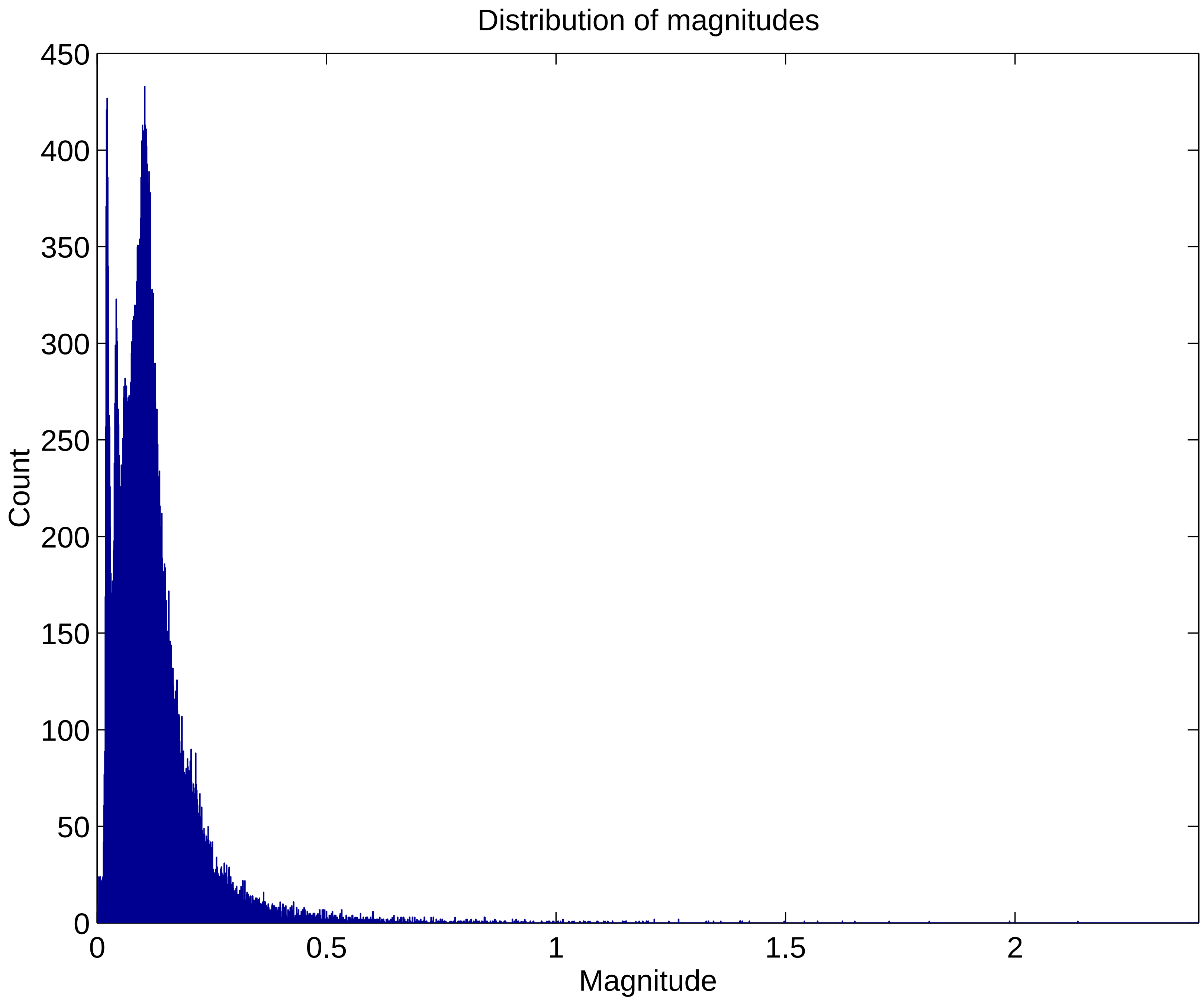
- Types: Left geometric, Right geometric, Uniform
- Magnitude ⇒ Left geometric
- Phase ⇒ Uniform

Left Geometric Distribution



Uniform Distribution




Distribution of magnitudes
Count
450
400
350
300
250
200
150
100
50
0
0
0.5
1
1.5
2
Magnitude

Speed up : Early Termination

- Obtain enough (not all) information to rank documents
- Sort index by term, value
- Rank top N documents
- Stopping criteria
  - Quit
  - Continue
- Accumulator (buffer) of size d used to gather results of top d documents

Skip Example

Early Termination : Sort Index

	d_1	d_2	d_3	d_4	d_5
term 1	0	2	0	1	1
term 2	1	0	1	2	0
term 3	2	1	0	0	2
term 4	1	1	3	0	0

Early Termination : Sort Index

term 1	$\langle d_1, 0\rangle$	$\langle d_2, 2\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 1\rangle$	$\langle d_5, 1\rangle$
term 2	$\langle d_1, 1\rangle$	$\langle d_2, 0\rangle$	$\langle d_3, 1\rangle$	$\langle d_4, 2\rangle$	$\langle d_5, 0\rangle$
term 3	$\langle d_1, 2\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 2\rangle$
term 4	$\langle d_1, 1\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 3\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 0\rangle$

Early Termination : Sort Index

term 1	$\langle d_1, 0\rangle$	$\langle d_2, 2\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 1\rangle$	$\langle d_5, 1\rangle$
term 2	$\langle d_1, 1\rangle$	$\langle d_2, 0\rangle$	$\langle d_3, 1\rangle$	$\langle d_4, 2\rangle$	$\langle d_5, 0\rangle$
term 3	$\langle d_1, 2\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 2\rangle$
term 4	$\langle d_1, 1\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 3\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 0\rangle$

$$\Downarrow$$

term 1	$\langle d_2, 2\rangle$	$\langle d_4, 1\rangle$	$\langle d_5, 1\rangle$	$\langle d_1, 0\rangle$	$\langle d_3, 0\rangle$
term 2	$\langle d_4, 2\rangle$	$\langle d_1, 1\rangle$	$\langle d_3, 1\rangle$	$\langle d_2, 0\rangle$	$\langle d_5, 0\rangle$
term 3	$\langle d_1, 2\rangle$	$\langle d_5, 2\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 0\rangle$
term 4	$\langle d_3, 3\rangle$	$\langle d_1, 1\rangle$	$\langle d_2, 1\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 0\rangle$

Early Termination : *Quit* Example

Query : spider car, Accumulator size = 3

Doc	Score
-	0
-	0
-	0

$\Leftarrow$

spider	$\langle d_2, 2\rangle$	$\langle d_4, 1\rangle$	$\langle d_5, 1\rangle$	$\langle d_1, 0\rangle$	$\langle d_3, 0\rangle$
house	$\langle d_4, 2\rangle$	$\langle d_1, 1\rangle$	$\langle d_3, 1\rangle$	$\langle d_2, 0\rangle$	$\langle d_5, 0\rangle$
car	$\langle d_1, 2\rangle$	$\langle d_5, 2\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 0\rangle$
boat	$\langle d_3, 3\rangle$	$\langle d_1, 1\rangle$	$\langle d_2, 1\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 0\rangle$

Early Termination : *Quit* Example

Query : spider car, Accumulator size = 3

Doc	Score
-	0
-	0
-	0

$\Leftarrow$

spider	$\langle d_2, 2\rangle$	$\langle d_4, 1\rangle$	$\langle d_5, 1\rangle$	$\langle d_1, 0\rangle$	$\langle d_3, 0\rangle$
house	$\langle d_4, 2\rangle$	$\langle d_1, 1\rangle$	$\langle d_3, 1\rangle$	$\langle d_2, 0\rangle$	$\langle d_5, 0\rangle$
car	$\langle d_1, 2\rangle$	$\langle d_5, 2\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 0\rangle$
boat	$\langle d_3, 3\rangle$	$\langle d_1, 1\rangle$	$\langle d_2, 1\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 0\rangle$

Early Termination : *Quit* Example

Query : spider car, Accumulator size = 3

Doc	Score
-	0
-	0
-	0

$\Leftarrow$

spider	$\langle d_2, 2\rangle$	$\langle d_4, 1\rangle$	$\langle d_5, 1\rangle$	$\langle d_1, 0\rangle$	$\langle d_3, 0\rangle$
house	$\langle d_4, 2\rangle$	$\langle d_1, 1\rangle$	$\langle d_3, 1\rangle$	$\langle d_2, 0\rangle$	$\langle d_5, 0\rangle$
car	$\langle d_1, 2\rangle$	$\langle d_5, 2\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 0\rangle$
boat	$\langle d_3, 3\rangle$	$\langle d_1, 1\rangle$	$\langle d_2, 1\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 0\rangle$

Early Termination : *Quit* Example

Query : spider car, Accumulator size = 3

Doc	Score
-	0
-	0
-	0

$\Leftarrow$

spider	$\langle d_2, 2\rangle$	$\langle d_4, 1\rangle$	$\langle d_5, 1\rangle$	$\langle d_1, 0\rangle$	$\langle d_3, 0\rangle$
house	$\langle d_4, 2\rangle$	$\langle d_1, 1\rangle$	$\langle d_3, 1\rangle$	$\langle d_2, 0\rangle$	$\langle d_5, 0\rangle$
car	$\langle d_1, 2\rangle$	$\langle d_5, 2\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 0\rangle$
boat	$\langle d_3, 3\rangle$	$\langle d_1, 1\rangle$	$\langle d_2, 1\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 0\rangle$

Early Termination : *Quit* Example

Query : spider car, Accumulator size = 3

Doc	Score
d_2	2
-	0
-	0

$\Leftarrow$

spider	$\langle d_2, 2\rangle$	$\langle d_4, 1\rangle$	$\langle d_5, 1\rangle$	$\langle d_1, 0\rangle$	$\langle d_3, 0\rangle$
house	$\langle d_4, 2\rangle$	$\langle d_1, 1\rangle$	$\langle d_3, 1\rangle$	$\langle d_2, 0\rangle$	$\langle d_5, 0\rangle$
car	$\langle d_1, 2\rangle$	$\langle d_5, 2\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 0\rangle$
boat	$\langle d_3, 3\rangle$	$\langle d_1, 1\rangle$	$\langle d_2, 1\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 0\rangle$

Early Termination : *Quit* Example

Query : spider car, Accumulator size = 3

Doc	Score
d_2	2
-	0
-	0

$\Leftarrow$

spider	$\langle d_2, 2\rangle$	$\langle d_4, 1\rangle$	$\langle d_5, 1\rangle$	$\langle d_1, 0\rangle$	$\langle d_3, 0\rangle$
house	$\langle d_4, 2\rangle$	$\langle d_1, 1\rangle$	$\langle d_3, 1\rangle$	$\langle d_2, 0\rangle$	$\langle d_5, 0\rangle$
car	$\langle d_1, 2\rangle$	$\langle d_5, 2\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 0\rangle$
boat	$\langle d_3, 3\rangle$	$\langle d_1, 1\rangle$	$\langle d_2, 1\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 0\rangle$

Early Termination : *Quit* Example

Query : spider car, Accumulator size = 3

Doc	Score
d_2	2
d_1	2
-	0

$\Leftarrow$

spider	$\langle d_2, 2\rangle$	$\langle d_4, 1\rangle$	$\langle d_5, 1\rangle$	$\langle d_1, 0\rangle$	$\langle d_3, 0\rangle$
house	$\langle d_4, 2\rangle$	$\langle d_1, 1\rangle$	$\langle d_3, 1\rangle$	$\langle d_2, 0\rangle$	$\langle d_5, 0\rangle$
car	$\langle d_1, 2\rangle$	$\langle d_5, 2\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 0\rangle$
boat	$\langle d_3, 3\rangle$	$\langle d_1, 1\rangle$	$\langle d_2, 1\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 0\rangle$

Early Termination : *Quit* Example

Query : spider car, Accumulator size = 3

Doc	Score
d_2	2
d_1	2
-	0

$\Leftarrow$

spider	$\langle d_2, 2\rangle$	$\langle d_4, 1\rangle$	$\langle d_5, 1\rangle$	$\langle d_1, 0\rangle$	$\langle d_3, 0\rangle$
house	$\langle d_4, 2\rangle$	$\langle d_1, 1\rangle$	$\langle d_3, 1\rangle$	$\langle d_2, 0\rangle$	$\langle d_5, 0\rangle$
car	$\langle d_1, 2\rangle$	$\langle d_5, 2\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 0\rangle$
boat	$\langle d_3, 3\rangle$	$\langle d_1, 1\rangle$	$\langle d_2, 1\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 0\rangle$

Early Termination : *Quit* Example

Query : spider car, Accumulator size = 3

Doc	Score
d_2	2
d_1	2
d_5	2

$\Leftarrow$

spider	$\langle d_2, 2\rangle$	$\langle d_4, 1\rangle$	$\langle d_5, 1\rangle$	$\langle d_1, 0\rangle$	$\langle d_3, 0\rangle$
house	$\langle d_4, 2\rangle$	$\langle d_1, 1\rangle$	$\langle d_3, 1\rangle$	$\langle d_2, 0\rangle$	$\langle d_5, 0\rangle$
car	$\langle d_1, 2\rangle$	$\langle d_5, 2\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 0\rangle$
boat	$\langle d_3, 3\rangle$	$\langle d_1, 1\rangle$	$\langle d_2, 1\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 0\rangle$

Early Termination : *Quit* Example

Query : spider car, Accumulator size = 3

Doc	Score
d_2	2
d_1	2
d_5	2

$\Leftarrow$

spider	$\langle d_2, 2\rangle$	$\langle d_4, 1\rangle$	$\langle d_5, 1\rangle$	$\langle d_1, 0\rangle$	$\langle d_3, 0\rangle$
house	$\langle d_4, 2\rangle$	$\langle d_1, 1\rangle$	$\langle d_3, 1\rangle$	$\langle d_2, 0\rangle$	$\langle d_5, 0\rangle$
car	$\langle d_1, 2\rangle$	$\langle d_5, 2\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 0\rangle$
boat	$\langle d_3, 3\rangle$	$\langle d_1, 1\rangle$	$\langle d_2, 1\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 0\rangle$

$\Downarrow$

Rank	Document
1	d_2
2	d_1
3	d_5

Early Termination : *Continue* Example

Query : spider car, Accumulator size = 3

Doc	Score
d_2	2
d_1	2
d_5	2

$\Leftarrow$

spider	$\langle d_2, 2\rangle$	$\langle d_4, 1\rangle$	$\langle d_5, 1\rangle$	$\langle d_1, 0\rangle$	$\langle d_3, 0\rangle$
house	$\langle d_4, 2\rangle$	$\langle d_1, 1\rangle$	$\langle d_3, 1\rangle$	$\langle d_2, 0\rangle$	$\langle d_5, 0\rangle$
car	$\langle d_1, 2\rangle$	$\langle d_5, 2\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 0\rangle$
boat	$\langle d_3, 3\rangle$	$\langle d_1, 1\rangle$	$\langle d_2, 1\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 0\rangle$

Early Termination : *Continue* Example

Query : spider car, Accumulator size = 3

Doc	Score
d_2	2
d_1	2
d_5	2

$\Leftarrow$

spider	$\langle d_2, 2\rangle$	$\langle d_4, 1\rangle$	$\langle d_5, 1\rangle$	$\langle d_1, 0\rangle$	$\langle d_3, 0\rangle$
house	$\langle d_4, 2\rangle$	$\langle d_1, 1\rangle$	$\langle d_3, 1\rangle$	$\langle d_2, 0\rangle$	$\langle d_5, 0\rangle$
car	$\langle d_1, 2\rangle$	$\langle d_5, 2\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 0\rangle$
boat	$\langle d_3, 3\rangle$	$\langle d_1, 1\rangle$	$\langle d_2, 1\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 0\rangle$

Early Termination : *Continue* Example

Query : spider car, Accumulator size = 3

Doc	Score
d_2	2
d_1	2
d_5	3

$\Leftarrow$

spider	$\langle d_2, 2\rangle$	$\langle d_4, 1\rangle$	$\langle d_5, 1\rangle$	$\langle d_1, 0\rangle$	$\langle d_3, 0\rangle$
house	$\langle d_4, 2\rangle$	$\langle d_1, 1\rangle$	$\langle d_3, 1\rangle$	$\langle d_2, 0\rangle$	$\langle d_5, 0\rangle$
car	$\langle d_1, 2\rangle$	$\langle d_5, 2\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 0\rangle$
boat	$\langle d_3, 3\rangle$	$\langle d_1, 1\rangle$	$\langle d_2, 1\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 0\rangle$

Early Termination : *Continue* Example

Query : spider car, Accumulator size = 3

Doc	Score
d_2	3
d_1	2
d_5	3

$\Leftarrow$

spider	$\langle d_2, 2\rangle$	$\langle d_4, 1\rangle$	$\langle d_5, 1\rangle$	$\langle d_1, 0\rangle$	$\langle d_3, 0\rangle$
house	$\langle d_4, 2\rangle$	$\langle d_1, 1\rangle$	$\langle d_3, 1\rangle$	$\langle d_2, 0\rangle$	$\langle d_5, 0\rangle$
car	$\langle d_1, 2\rangle$	$\langle d_5, 2\rangle$	$\langle d_2, 1\rangle$	$\langle d_3, 0\rangle$	$\langle d_4, 0\rangle$
boat	$\langle d_3, 3\rangle$	$\langle d_1, 1\rangle$	$\langle d_2, 1\rangle$	$\langle d_4, 0\rangle$	$\langle d_5, 0\rangle$

$\Downarrow$

Rank	Document
1	d_2
2	d_5
3	d_1

Spectral Early Termination

- Sort data by term, bin, magnitude
- Must combine magnitude and phase across terms before combining
- Need Magnitude and phase accumulators for each bin.

Skip Example

Spectral Early Termination : Sort Index

term 1	d_1	[0.3	-2.1	0.4	-1.1]
term 1	d_2	[0.7	0.4	0	-1.5]
term 1	d_3	[1.2	2.1	-0.3	0.5]
term 1	d_4	[0	0	0	0]
term 2	d_1	[1.9	-1.1	0.6	1.2]
term 2	d_2	[0	0	0	0]
term 2	d_3	[0.9	1.7	-1.1	-1.4]
term 2	d_4	[1.3	0.7	-0.8	1.1]

Spectral Early Termination : Sort Index

term 1	$\langle d_1, 1, 0.3\rangle$	$\langle d_1, 2, -2.1\rangle$	$\langle d_1, 3, 0.4\rangle$	$\langle d_1, 4, -1.1\rangle$
term 1	$\langle d_2, 1, 0.7\rangle$	$\langle d_2, 2, 0.4\rangle$	$\langle d_2, 3, 0\rangle$	$\langle d_2, 4, -1.5\rangle$
term 1	$\langle d_3, 1, 1.2\rangle$	$\langle d_3, 2, 2.1\rangle$	$\langle d_3, 3, -0.3\rangle$	$\langle d_3, 4, 0.5\rangle$
term 1	$\langle d_4, 1, 0\rangle$	$\langle d_4, 2, 0\rangle$	$\langle d_4, 3, 0\rangle$	$\langle d_4, 4, 0\rangle$
term 2	$\langle d_1, 1, 1.9\rangle$	$\langle d_1, 2, -1.1\rangle$	$\langle d_1, 3, 0.6\rangle$	$\langle d_1, 4, 1.2\rangle$
term 2	$\langle d_2, 1, 0\rangle$	$\langle d_2, 2, 0\rangle$	$\langle d_2, 3, 0\rangle$	$\langle d_2, 4, 0\rangle$
term 2	$\langle d_3, 1, 0.9\rangle$	$\langle d_3, 2, 1.7\rangle$	$\langle d_3, 3, -1.1\rangle$	$\langle d_3, 4, -1.4\rangle$
term 2	$\langle d_4, 1, 1.3\rangle$	$\langle d_4, 2, 0.7\rangle$	$\langle d_4, 3, -0.8\rangle$	$\langle d_4, 4, 1.1\rangle$

Spectral Early Termination : Sort Index

term 1	$\langle d_1, 1, 0.3\rangle$	$\langle d_1, 2, -2.1\rangle$	$\langle d_1, 3, 0.4\rangle$	$\langle d_1, 4, -1.1\rangle$
term 1	$\langle d_2, 1, 0.7\rangle$	$\langle d_2, 2, 0.4\rangle$	$\langle d_2, 3, 0\rangle$	$\langle d_2, 4, -1.5\rangle$
term 1	$\langle d_3, 1, 1.2\rangle$	$\langle d_3, 2, 2.1\rangle$	$\langle d_3, 3, -0.3\rangle$	$\langle d_3, 4, 0.5\rangle$
term 1	$\langle d_4, 1, 0\rangle$	$\langle d_4, 2, 0\rangle$	$\langle d_4, 3, 0\rangle$	$\langle d_4, 4, 0\rangle$
term 2	$\langle d_1, 1, 1.9\rangle$	$\langle d_1, 2, -1.1\rangle$	$\langle d_1, 3, 0.6\rangle$	$\langle d_1, 4, 1.2\rangle$
term 2	$\langle d_2, 1, 0\rangle$	$\langle d_2, 2, 0\rangle$	$\langle d_2, 3, 0\rangle$	$\langle d_2, 4, 0\rangle$
term 2	$\langle d_3, 1, 0.9\rangle$	$\langle d_3, 2, 1.7\rangle$	$\langle d_3, 3, -1.1\rangle$	$\langle d_3, 4, -1.4\rangle$
term 2	$\langle d_4, 1, 1.3\rangle$	$\langle d_4, 2, 0.7\rangle$	$\langle d_4, 3, -0.8\rangle$	$\langle d_4, 4, 1.1\rangle$

Spectral Early Termination : Sort Index

term 1	$\langle d_3, 1, 1.2\rangle$	$\langle d_2, 1, 0.7\rangle$	$\langle d_1, 1, 0.3\rangle$	$\langle d_1, 2, -2.1\rangle$
	$\langle d_3, 2, 2.1\rangle$	$\langle d_2, 2, 0.4\rangle$	$\langle d_1, 3, 0.4\rangle$	$\langle d_3, 3, -0.3\rangle$
	$\langle d_2, 4, -1.5\rangle$	$\langle d_1, 4, -1.1\rangle$	$\langle d_3, 4, 0.5\rangle$	
term 2	$\langle d_1, 1, 1.9\rangle$	$\langle d_4, 1, 1.3\rangle$	$\langle d_3, 1, 0.9\rangle$	$\langle d_3, 2, 1.7\rangle$
	$\langle d_1, 2, -1.1\rangle$	$\langle d_4, 2, 0.7\rangle$	$\langle d_3, 3, -1.1\rangle$	$\langle d_4, 3, -0.8\rangle$
	$\langle d_1, 3, 0.6\rangle$	$\langle d_3, 4, -1.4\rangle$	$\langle d_1, 4, 1.2\rangle$	$\langle d_4, 4, 1.1\rangle$

Early Termination : *Quit-n* Example

Query : spider car, Accumulator size = 3, 2 bins

Doc	Magnitude	
	Bin 1	Bin 2
-	0	0
-	0	0
-	0	0

Doc	Phase	
	Bin 1	Bin 2
-	0	0
-	0	0
-	0	0

$\Uparrow$

spider	$\langle d_3, 1, 1.2 \rangle$	$\langle d_2, 1, 0.7 \rangle$	$\langle d_1, 1, 0.3 \rangle$	$\langle d_1, 2, -2.1 \rangle$
	$\langle d_3, 2, 2.1 \rangle$	$\langle d_2, 2, 0.4 \rangle$	$\langle d_1, 3, 0.4 \rangle$	$\langle d_3, 3, -0.3 \rangle$
car	$\langle d_1, 1, 1.9 \rangle$	$\langle d_4, 1, 1.3 \rangle$	$\langle d_3, 1, 0.9 \rangle$	$\langle d_3, 2, 1.7 \rangle$
	$\langle d_1, 2, -1.1 \rangle$	$\langle d_4, 2, 0.7 \rangle$	$\langle d_3, 3, -1.1 \rangle$	$\langle d_4, 3, -0.8 \rangle$

Early Termination : *Quit-n* Example

Query : spider car, Accumulator size = 3, 2 bins

Doc	Magnitude	
	Bin 1	Bin 2
-	0	0
-	0	0
-	0	0

Doc	Phase	
	Bin 1	Bin 2
-	0	0
-	0	0
-	0	0

$\Uparrow$

spider	$\langle d_3, 1, 1.2\rangle$	$\langle d_2, 1, 0.7\rangle$	$\langle d_1, 1, 0.3\rangle$	$\langle d_1, 2, -2.1\rangle$
	$\langle d_3, 2, 2.1\rangle$	$\langle d_2, 2, 0.4\rangle$	$\langle d_1, 3, 0.4\rangle$	$\langle d_3, 3, -0.3\rangle$
car	$\langle d_1, 1, 1.9\rangle$	$\langle d_4, 1, 1.3\rangle$	$\langle d_3, 1, 0.9\rangle$	$\langle d_3, 2, 1.7\rangle$
	$\langle d_1, 2, -1.1\rangle$	$\langle d_4, 2, 0.7\rangle$	$\langle d_3, 3, -1.1\rangle$	$\langle d_4, 3, -0.8\rangle$

Early Termination : *Quit-n* Example

Query : spider car, Accumulator size = 3, 2 bins

Doc	Magnitude	
	Bin 1	Bin 2
d_1	1.9	0
-	0	0
-	0	0

Doc	Phase	
	Bin 1	Bin 2
d_1	1	0
-	0	0
-	0	0

$\Uparrow$

spider	$\langle d_3, 1, 1.2\rangle$	$\langle d_2, 1, 0.7\rangle$	$\langle d_1, 1, 0.3\rangle$	$\langle d_1, 2, -2.1\rangle$
	$\langle d_3, 2, 2.1\rangle$	$\langle d_2, 2, 0.4\rangle$	$\langle d_1, 3, 0.4\rangle$	$\langle d_3, 3, -0.3\rangle$
car	$\langle d_1, 1, 1.9\rangle$	$\langle d_4, 1, 1.3\rangle$	$\langle d_3, 1, 0.9\rangle$	$\langle d_3, 2, 1.7\rangle$
	$\langle d_1, 2, -1.1\rangle$	$\langle d_4, 2, 0.7\rangle$	$\langle d_3, 3, -1.1\rangle$	$\langle d_4, 3, -0.8\rangle$

Early Termination : *Quit-n* Example

Query : spider car, Accumulator size = 3, 2 bins

Doc	Magnitude	
	Bin 1	Bin 2
d_1	1.9	0
-	0	0
-	0	0

Doc	Phase	
	Bin 1	Bin 2
d_1	1	0
-	0	0
-	0	0

$\Uparrow$

spider	$\langle d_3, 1, 1.2 \rangle$	$\langle d_2, 1, 0.7 \rangle$	$\langle d_1, 1, 0.3 \rangle$	$\langle d_1, 2, -2.1 \rangle$
	$\langle d_3, 2, 2.1 \rangle$	$\langle d_2, 2, 0.4 \rangle$	$\langle d_1, 3, 0.4 \rangle$	$\langle d_3, 3, -0.3 \rangle$
car	$\langle d_1, 1, 1.9 \rangle$	$\langle d_4, 1, 1.3 \rangle$	$\langle d_3, 1, 0.9 \rangle$	$\langle d_3, 2, 1.7 \rangle$
	$\langle d_1, 2, -1.1 \rangle$	$\langle d_4, 2, 0.7 \rangle$	$\langle d_3, 3, -1.1 \rangle$	$\langle d_4, 3, -0.8 \rangle$

CUBIN

Early Termination : *Quit-n* Example

Query : spider car, Accumulator size = 3, 2 bins

Doc	Magnitude	
	Bin 1	Bin 2
d_1	1.9	0
d_4	1.3	0
-	0	0

Doc	Phase	
	Bin 1	Bin 2
d_1	1	0
d_4	1	0
-	0	0

$\Uparrow$

spider	$\langle d_3, 1, 1.2 \rangle$	$\langle d_2, 1, 0.7 \rangle$	$\langle d_1, 1, 0.3 \rangle$	$\langle d_1, 2, -2.1 \rangle$
	$\langle d_3, 2, 2.1 \rangle$	$\langle d_2, 2, 0.4 \rangle$	$\langle d_1, 3, 0.4 \rangle$	$\langle d_3, 3, -0.3 \rangle$
car	$\langle d_1, 1, 1.9 \rangle$	$\langle d_4, 1, 1.3 \rangle$	$\langle d_3, 1, 0.9 \rangle$	$\langle d_3, 2, 1.7 \rangle$
	$\langle d_1, 2, -1.1 \rangle$	$\langle d_4, 2, 0.7 \rangle$	$\langle d_3, 3, -1.1 \rangle$	$\langle d_4, 3, -0.8 \rangle$

Early Termination : *Quit-n* Example

Query : spider car, Accumulator size = 3, 2 bins

Doc	Magnitude	
	Bin 1	Bin 2
d_1	1.9	0
d_4	1.3	0
-	0	0

Doc	Phase	
	Bin 1	Bin 2
d_1	1	0
d_4	1	0
-	0	0

$\Uparrow$

spider	$\langle d_3, 1, 1.2 \rangle$	$\langle d_2, 1, 0.7 \rangle$	$\langle d_1, 1, 0.3 \rangle$	$\langle d_1, 2, -2.1 \rangle$
	$\langle d_3, 2, 2.1 \rangle$	$\langle d_2, 2, 0.4 \rangle$	$\langle d_1, 3, 0.4 \rangle$	$\langle d_3, 3, -0.3 \rangle$
car	$\langle d_1, 1, 1.9 \rangle$	$\langle d_4, 1, 1.3 \rangle$	$\langle d_3, 1, 0.9 \rangle$	$\langle d_3, 2, 1.7 \rangle$
	$\langle d_1, 2, -1.1 \rangle$	$\langle d_4, 2, 0.7 \rangle$	$\langle d_3, 3, -1.1 \rangle$	$\langle d_4, 3, -0.8 \rangle$

Early Termination : *Quit-n* Example

Query : spider car, Accumulator size = 3, 2 bins

Doc	Magnitude	
	Bin 1	Bin 2
d_1	1.9	0
d_4	1.3	0
d_3	1.2	0

Doc	Phase	
	Bin 1	Bin 2
d_1	1	0
d_4	1	0
d_3	1	0

$\Uparrow$

spider	$\langle d_3, 1, 1.2\rangle$	$\langle d_2, 1, 0.7\rangle$	$\langle d_1, 1, 0.3\rangle$	$\langle d_1, 2, -2.1\rangle$
	$\langle d_3, 2, 2.1\rangle$	$\langle d_2, 2, 0.4\rangle$	$\langle d_1, 3, 0.4\rangle$	$\langle d_3, 3, -0.3\rangle$
car	$\langle d_1, 1, 1.9\rangle$	$\langle d_4, 1, 1.3\rangle$	$\langle d_3, 1, 0.9\rangle$	$\langle d_3, 2, 1.7\rangle$
	$\langle d_1, 2, -1.1\rangle$	$\langle d_4, 2, 0.7\rangle$	$\langle d_3, 3, -1.1\rangle$	$\langle d_4, 3, -0.8\rangle$

Early Termination : *Quit-n* Example

Query : spider car, Accumulator size = 3, 2 bins

Doc	Magnitude	
	Bin 1	Bin 2
d_1	1.9	0
d_4	1.3	0
d_3	1.2	0

Doc	Phase	
	Bin 1	Bin 2
d_1	1	0
d_4	1	0
d_3	1	0

$\Uparrow$

spider	$\langle d_3, 1, 1.2\rangle$	$\langle d_2, 1, 0.7\rangle$	$\langle d_1, 1, 0.3\rangle$	$\langle d_1, 2, -2.1\rangle$
	$\langle d_3, 2, 2.1\rangle$	$\langle d_2, 2, 0.4\rangle$	$\langle d_1, 3, 0.4\rangle$	$\langle d_3, 3, -0.3\rangle$
car	$\langle d_1, 1, 1.9\rangle$	$\langle d_4, 1, 1.3\rangle$	$\langle d_3, 1, 0.9\rangle$	$\langle d_3, 2, 1.7\rangle$
	$\langle d_1, 2, -1.1\rangle$	$\langle d_4, 2, 0.7\rangle$	$\langle d_3, 3, -1.1\rangle$	$\langle d_4, 3, -0.8\rangle$

Early Termination : *Quit-n* Example

Query : spider car, Accumulator size = 3, 2 bins

Doc	Magnitude	
	Bin 1	Bin 2
d_1	1.9	0
d_4	1.3	0
d_3	1.2	0

Doc	Phase	
	Bin 1	Bin 2
d_1	1	0
d_4	1	0
d_3	1	0

$\Uparrow$

spider	$\langle d_3, 1, 1.2\rangle$	$\langle d_2, 1, 0.7\rangle$	$\langle d_1, 1, 0.3\rangle$	$\langle d_1, 2, -2.1\rangle$
	$\langle d_3, 2, 2.1\rangle$	$\langle d_2, 2, 0.4\rangle$	$\langle d_1, 3, 0.4\rangle$	$\langle d_3, 3, -0.3\rangle$
car	$\langle d_1, 1, 1.9\rangle$	$\langle d_4, 1, 1.3\rangle$	$\langle d_3, 1, 0.9\rangle$	$\langle d_3, 2, 1.7\rangle$
	$\langle d_1, 2, -1.1\rangle$	$\langle d_4, 2, 0.7\rangle$	$\langle d_3, 3, -1.1\rangle$	$\langle d_4, 3, -0.8\rangle$

CUBIN

Early Termination : *Quit-n* Example

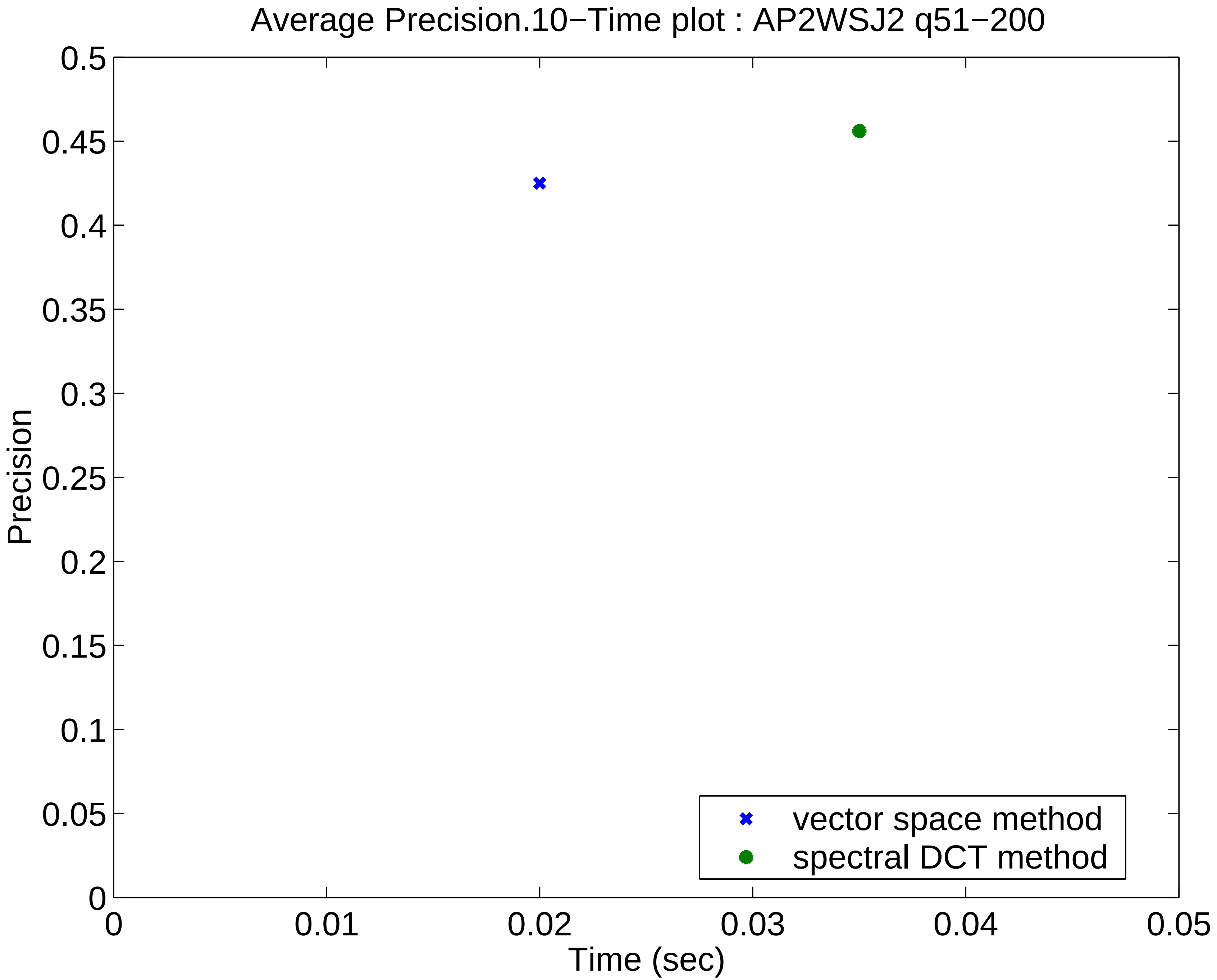
Query : spider car, Accumulator size = 3, 2 bins

Doc	Magnitude	
	Bin 1	Bin 2
d_1	2.2	3.2
d_4	1.3	0.7
d_3	2.1	3.8

Doc	Phase	
	Bin 1	Bin 2
d_1	2	-2
d_4	1	1
d_3	2	2

$\Uparrow$

spider	$\langle d_3, 1, 1.2\rangle$	$\langle d_2, 1, 0.7\rangle$	$\langle d_1, 1, 0.3\rangle$	$\langle d_1, 2, -2.1\rangle$
	$\langle d_3, 2, 2.1\rangle$	$\langle d_2, 2, 0.4\rangle$	$\langle d_1, 3, 0.4\rangle$	$\langle d_3, 3, -0.3\rangle$
car	$\langle d_1, 1, 1.9\rangle$	$\langle d_4, 1, 1.3\rangle$	$\langle d_3, 1, 0.9\rangle$	$\langle d_3, 2, 1.7\rangle$
	$\langle d_1, 2, -1.1\rangle$	$\langle d_4, 2, 0.7\rangle$	$\langle d_3, 3, -1.1\rangle$	$\langle d_4, 3, -0.8\rangle$


Average Precision.10−Time plot : AP2WSJ2 q51−200
Precision
Time (sec)
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0.01
0.02
0.03
0.04
0.05
vector space method
spectral DCT method

Outline

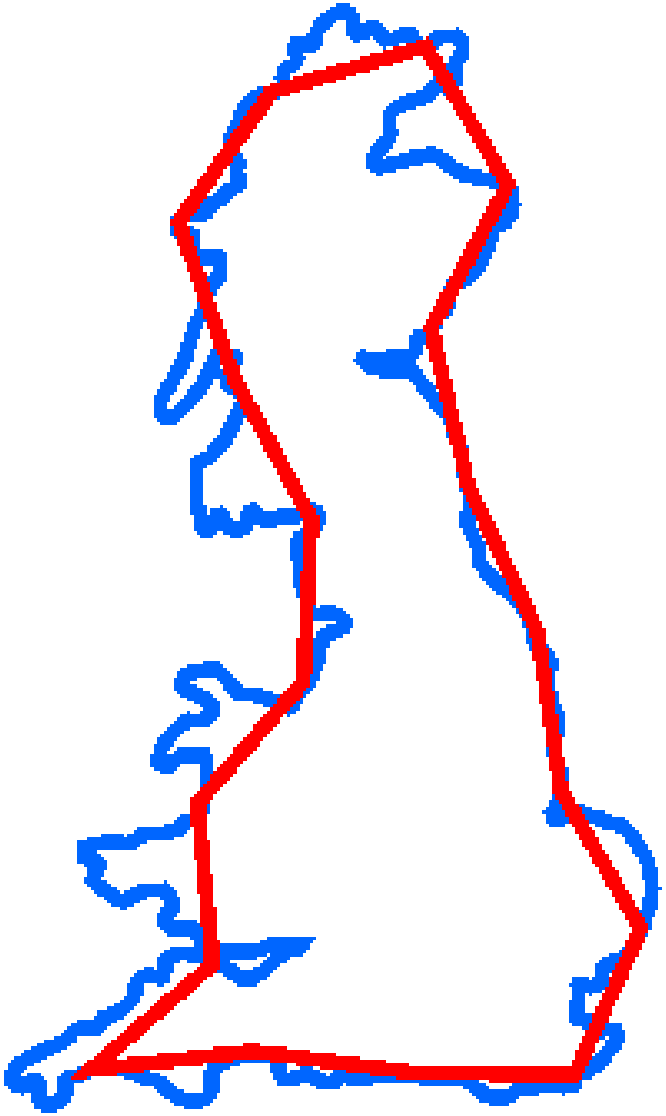
- Why do we need text retrieval?
- Our method : Spectral Document Ranking (SDR)
- Improving Efficiency of SDR
- Multiresolution Analysis with SDR
  - Document Resolution
  - Wavelet Decomposition
- Comparisons

Document Resolution

- Self similarity appears in nature everywhere
- Also appears in written language
- Mandelbrot : How long is the coast of Britain?
- Vector space methods = Big stick
- Fourier, Cosine, Gauss = Little stick
- Wavelet transform = Variable stick size

Document Resolution

- Self similarity appears in nature everywhere
- Also appears in written language
- Mandelbrot : How long is the coast of Britain?
- Vector space methods = Big stick
- Fourier, Cosine, Gauss = Little stick
- Wavelet transform = Variable stick size

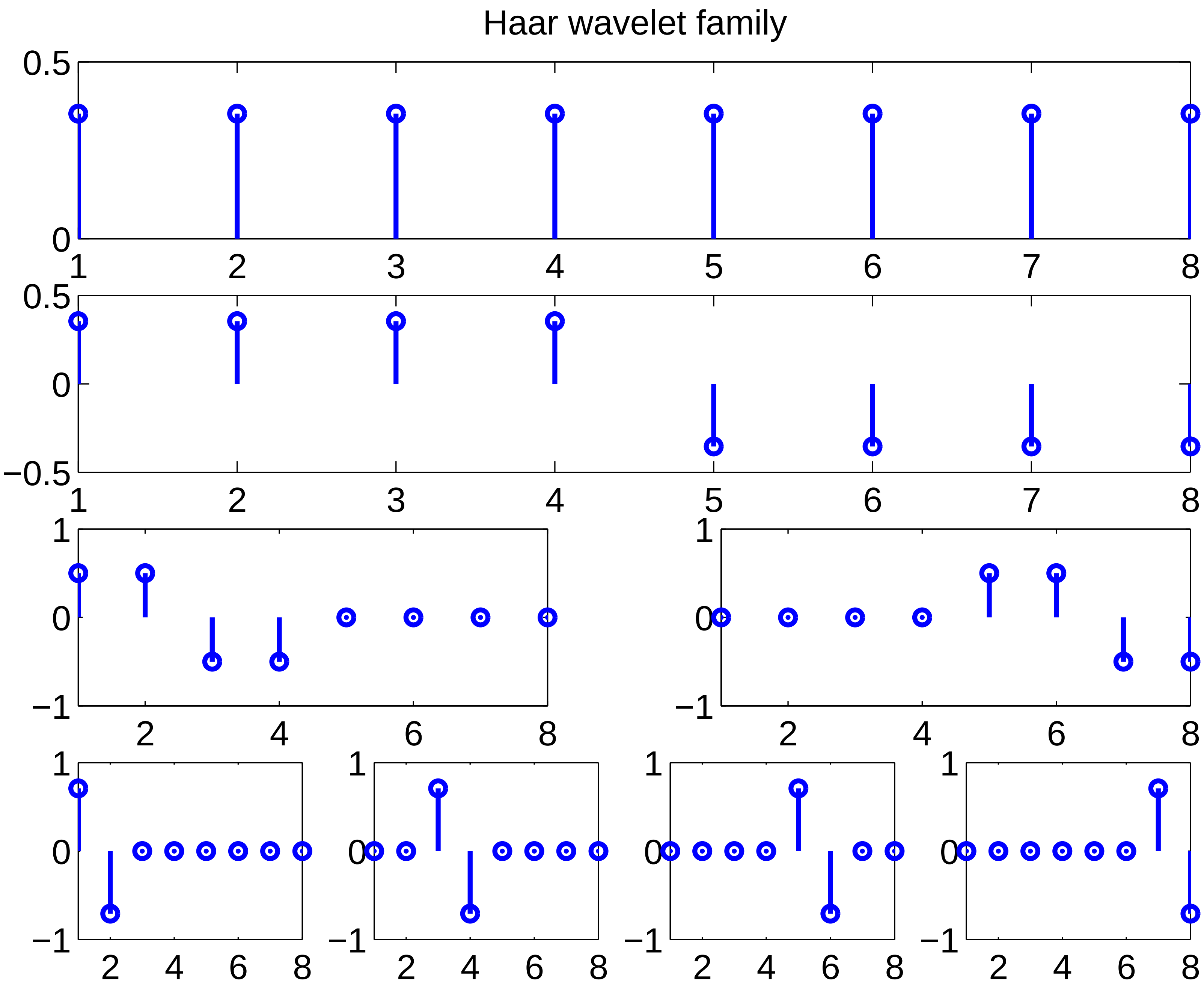


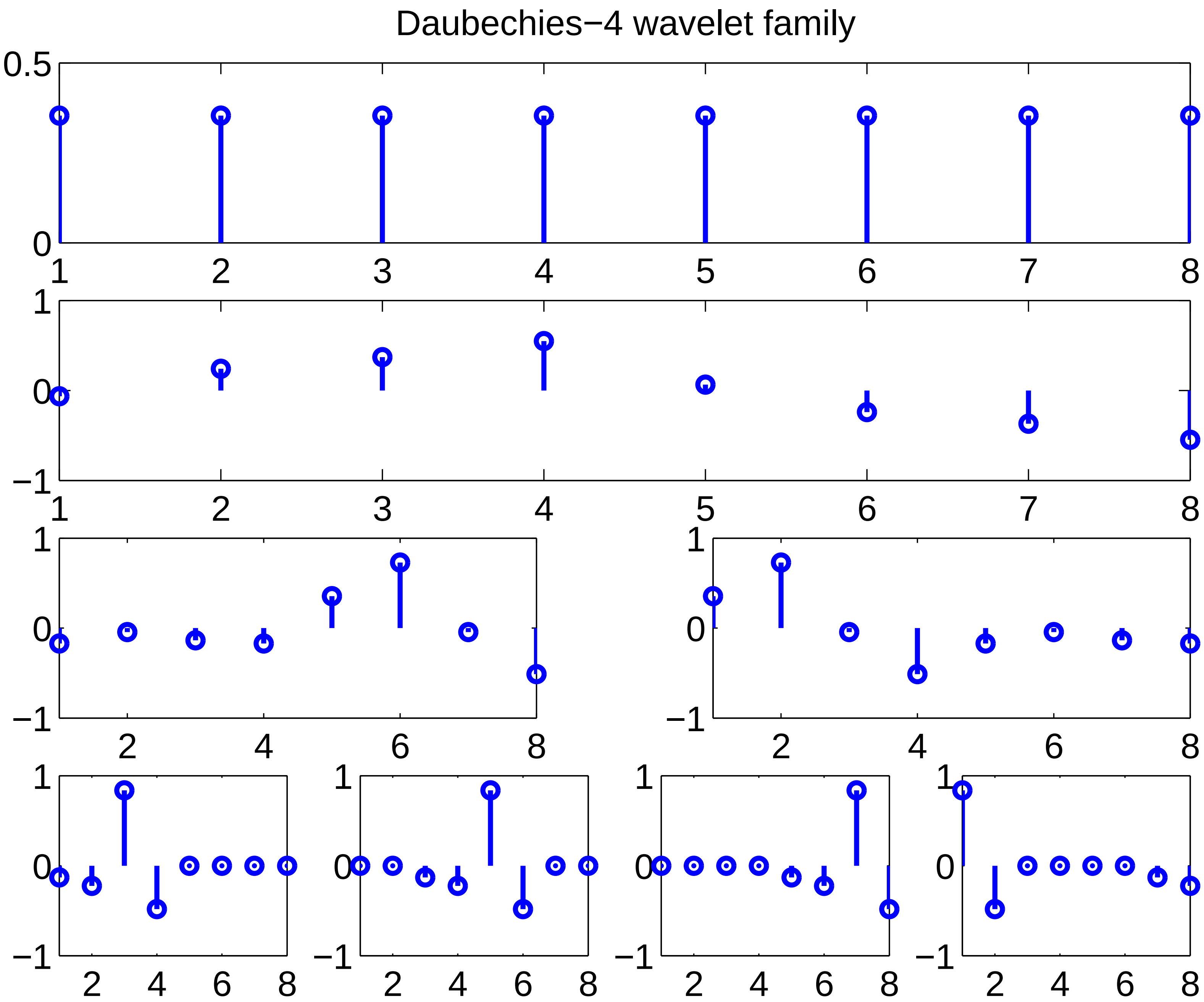
CUBIN

Wavelet Decomposition

- Signal is decomposed into orthogonal wavelet set.
- Wavelet set is a scaled and shifted version of the wavelet function
- Desired properties:
  - Small size of support is essential to focus on single term bins and less non-zero components (better compression)
- Wavelets chosen:
  - Haar wavelet
  - Daubechies-4 wavelet

Haar wavelet family





Daubechies−4 wavelet family
0.5
0
1
2
3
4
5
6
7
8
−1

Document Score

- Use same method as in Cosine transform method (eg. treat sign of spectral component as phase)
- Experimented with combining score components using:
  - Sum score components
  - Sum squared score components (to conserve signal energy)

Wavelet Precision Results

- Slightly better results than Fourier method
- Less components to store for better results
  - Fourier ≈ 10 components (5 real + 5 imag)
  - Haar ≈ 4 components
  - Daubechies-4 ≈ 6 components
- No complex calculations required (no imaginary values)

	fds-5-4-1	haar-5-4-6	daub4-5-4-6
Prec.5	0.4867	0.5000	0.4960
Prec.10	0.4647	0.4633	0.4653
Prec.15	0.4440	0.4404	0.4391
Prec.20	0.4193	0.4213	0.4177

Outline

- Why do we need text retrieval?
- Our method : Spectral Document Ranking (SDR)
- Improving Efficiency of SDR
- Multiresolution Analysis with SDR
- Comparisons

Comparison of Results

	VSM	Fourier	Cosine	Gauss	Haar	Daub-4
Speed	Fast	Slow	Fast	Slow	Fast	Fast
Storage	Avg	Large	Avg	Large	Avg	Avg
Prec.5	0.4440	0.4907	0.4907	0.4733	0.5000	0.4960
Prec.10	0.4247	0.4647	0.4547	0.4560	0.4633	0.4653
Prec.15	0.4142	0.4440	0.4369	0.4267	0.4404	0.4391
Prec.20	0.3953	0.4193	0.4147	0.4073	0.4213	0.4177

Future Directions

- Spectral document retrieval with feedback
  - Eg. using SVM's
- Develop document signals constructed from non-orthogonal term signals
- Topic extraction from documents
  - Eg. creating abstracts
- Spectral Audio (Speech or Music) Retrieval
  - Speech is easier to record than transcribe
  - Compressed speech is stored in its spectral form
  - The human ear sends frequency components to the brain
  - Analysing the audio spectrum is natural

CUBIN

Conclusions

- We have establised a need for a more precise text search technique
- We introduced new spectral document ranking methods (Fourier, Cosine, Wavelet)
- The Fourier transform method gave one of the best results, but was slow
- The Cosine transform method provided results similar to the Fourier method, but with faster query time and compact index.
- The Wavelet transform methods gave results just as good as the Fourier method, with a fast query time and compact index.