Beyond Text

INFM 718X/LBSC 708X

Session 10

Douglas W. Oard

Agenda

- Beyond Text, but still language
  - Scanned documents
  - Speech
- Beyond Text, but still information
  - Images
  - Video
- Beyond text to data

Elephant joke

From Wikipedia, the free encyclopedia

An **elephant joke** is a joke or riddle that involves an elephant. It usually relies on the great size and/or weight of the animal for its humor. Although elephant jokes are typically children's humor, a more sophisticated form appeals more to adults,

---

Elephant jokes are frequently nonsensical, and may in some cases be anti-jokes:

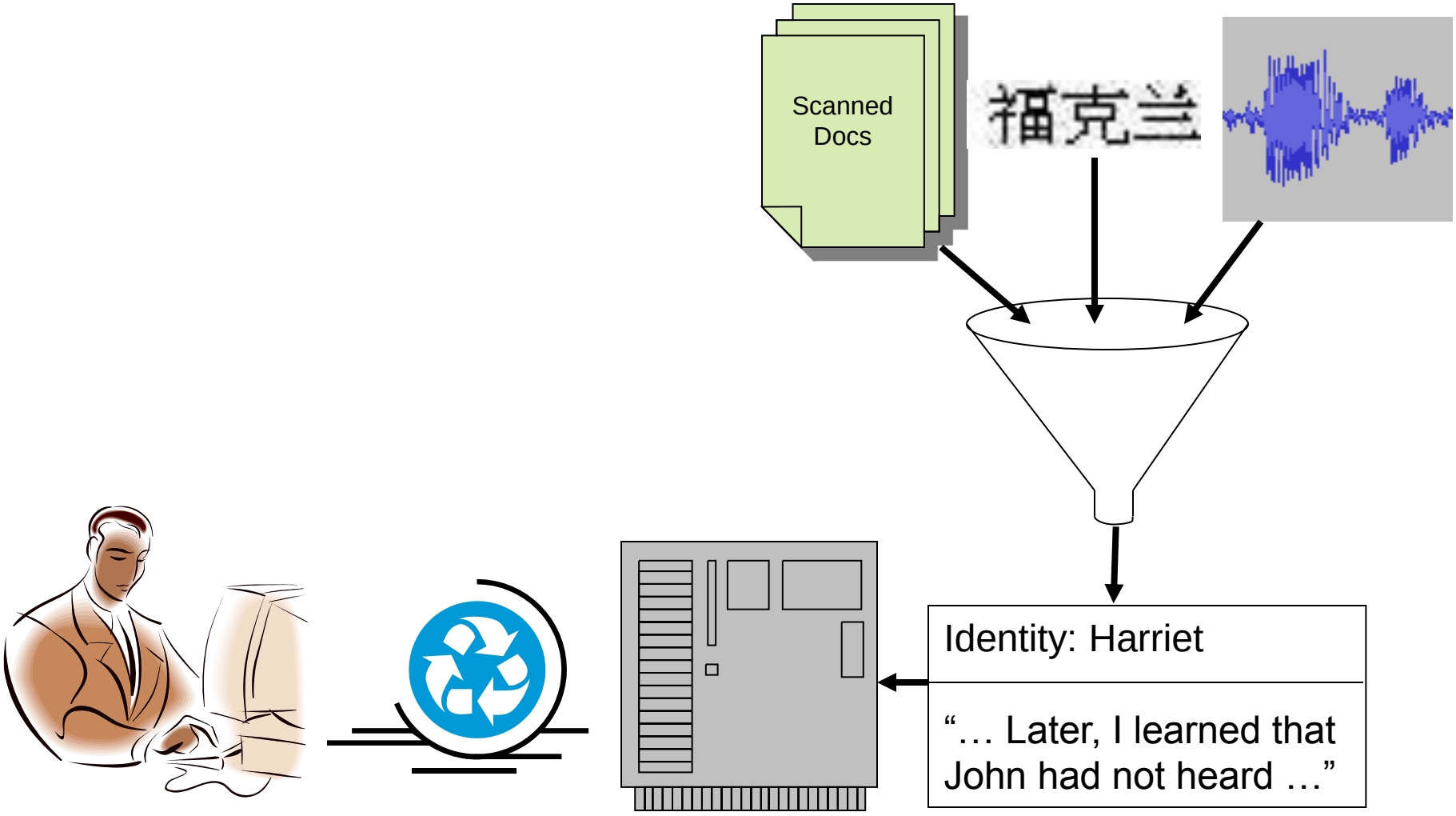
Q: How do you shoot a blue elephant?
A: With a blue elephant gun, of course.

Q: How do you shoot a yellow elephant?
A: Have you ever seen a yellow elephant?

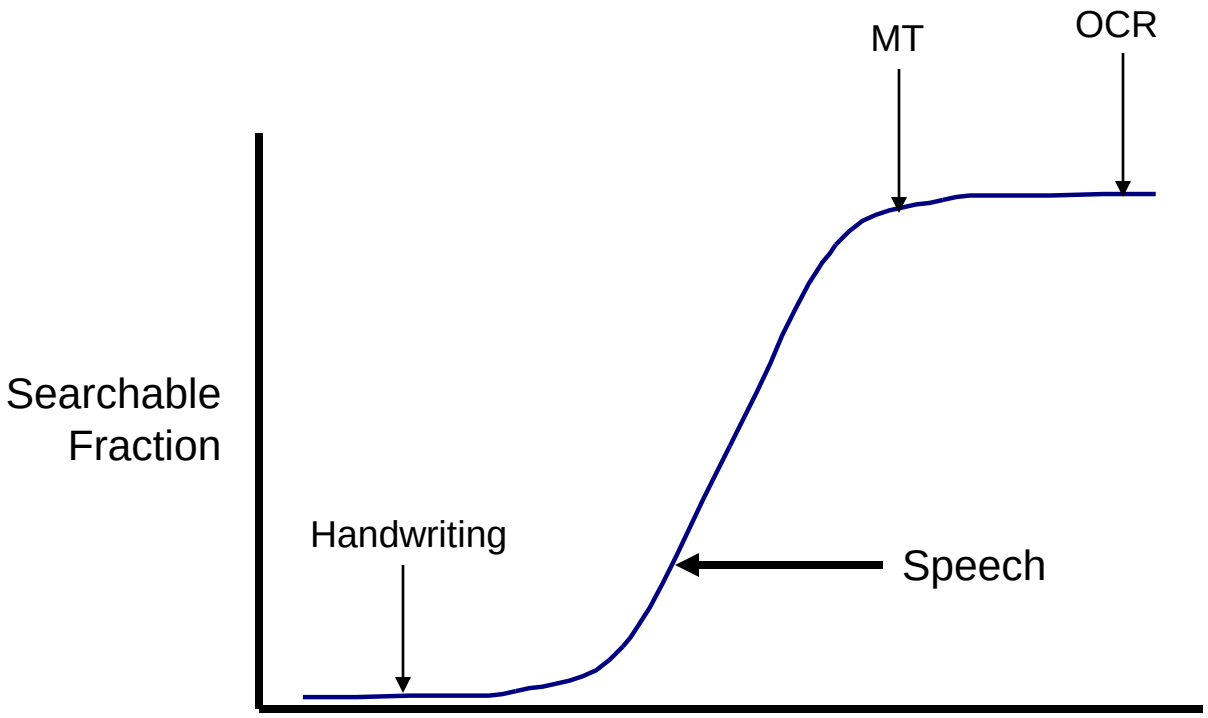
Q: How do you shoot a red elephant?
A: Hold his trunk shut until he turns blue, and then shoot him with the blue elephant gun.

Other standard variants

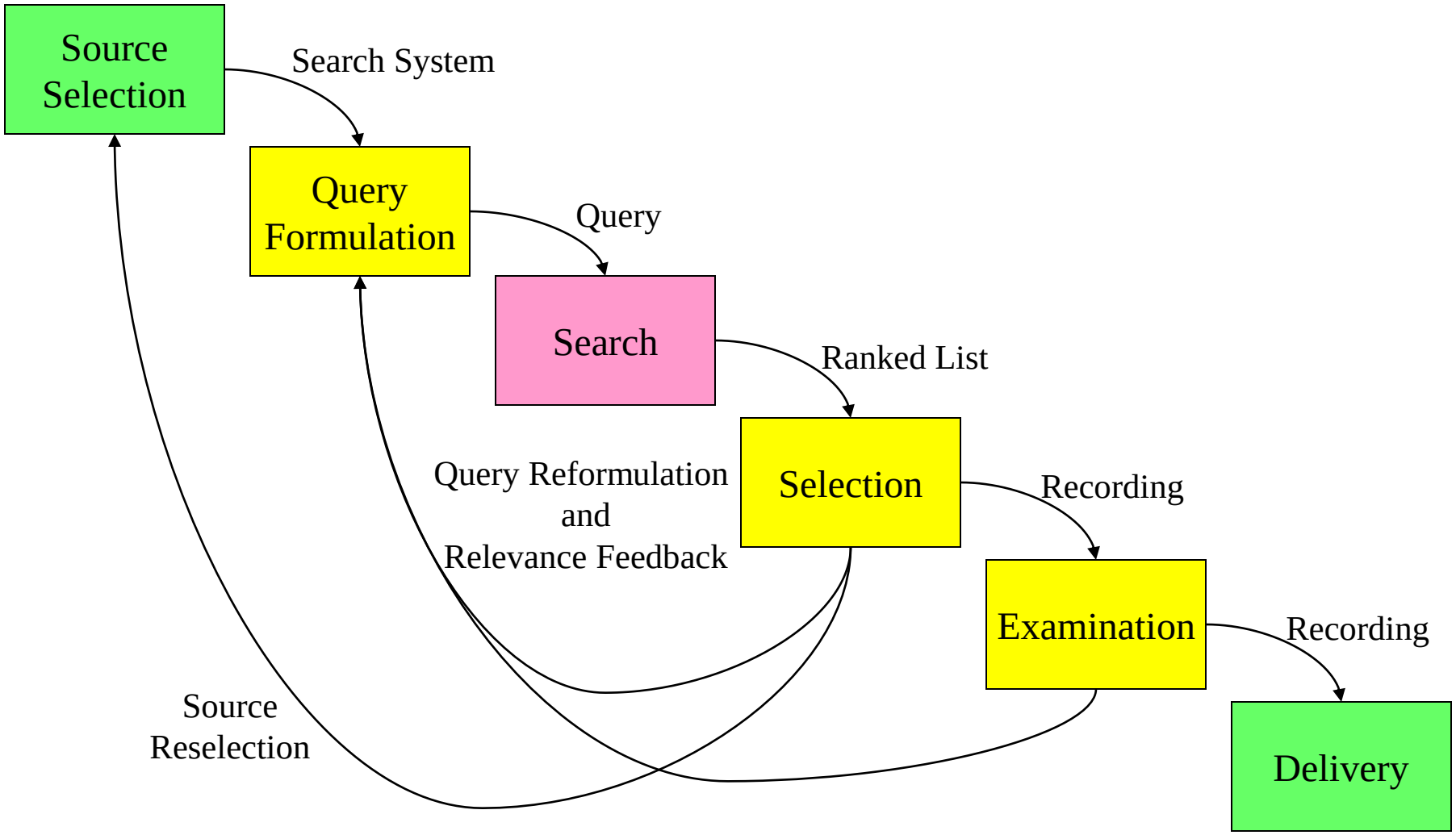
Expanding the Search Space



Language Conversion



Supporting Information Access



Indexing and Retrieving Images of Documents

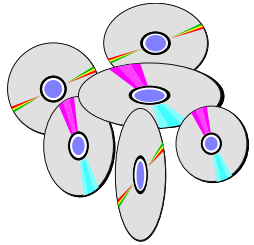
LBSC 796/INFM 718R

David Doermann, UMIACS

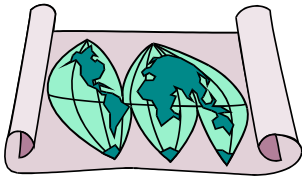
Sources

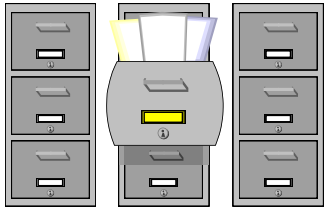
- Existing digitized collections
- Networked digital copiers
- High-volume digitization

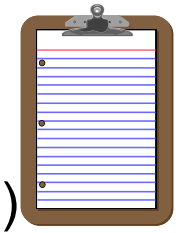
Document

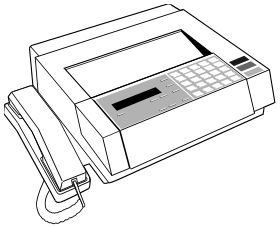


- Basic Medium for Recording Information
- Transient
  - Space
  - Time
- Multiple Forms
  - Hardcopy (paper, stone, ..) / Electronic (CDROM, Internet, …)
  - Written/Auditory/Visual (symbolic, scenic)
- Access Requirements
  - Search
  - Browse
  - “Read”









Images

- Pixel representation of intensity map
- No explicit “content”, only relations
- Image analysis
  - Attempts to mimic human visual behavior
  - Draw conclusions, hypothesize and verify

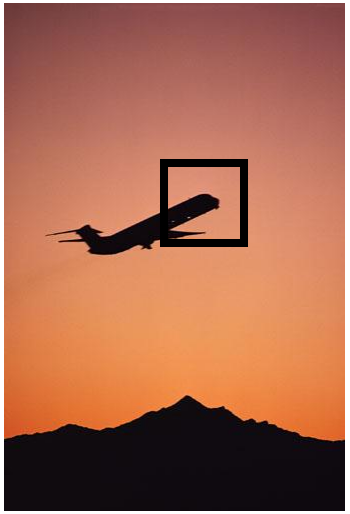
Image databases

Use primitive image analysis to represent content

Transform semantic queries into “image features”

color, shape, texture …

spatial relations



10	27	33	29
27	34	33	54
54	47	89	60
25	35	43	9

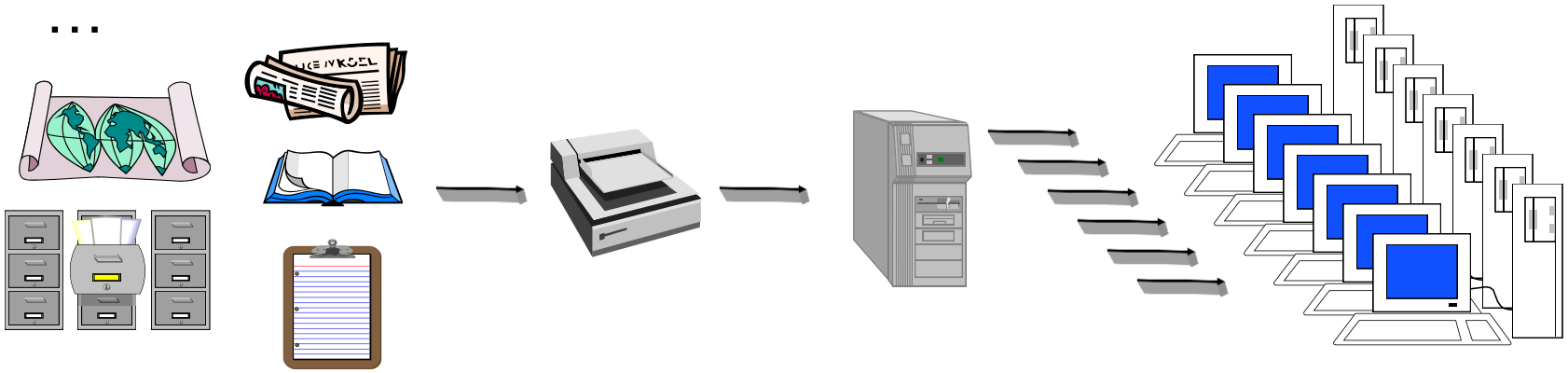
Document Images

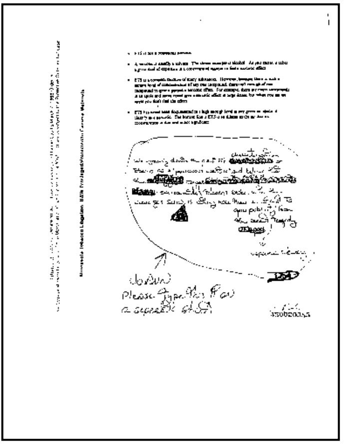
- A collection of dots called “pixels”
  - Arranged in a grid and called a “bitmap”
- Pixels often binary-valued (black, white)
  - But greyscale or color is sometimes needed
- 300 dots per inch (dpi) gives the best results
  - But images are quite large (1 MB per page)
  - Faxes are normally 72 dpi
- Usually stored in TIFF or PDF format

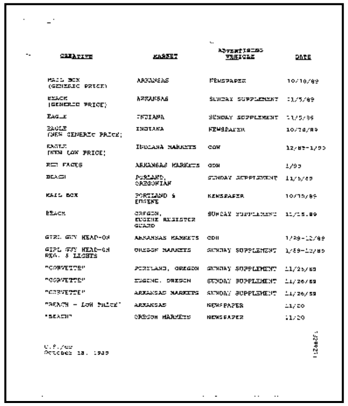
Yet we want to be able to process them like text files!

Document Image "Database"

- Collection of scanned images
- Need to be available for indexing and retrieval, abstracting, routing, editing, dissemination, interpretation …

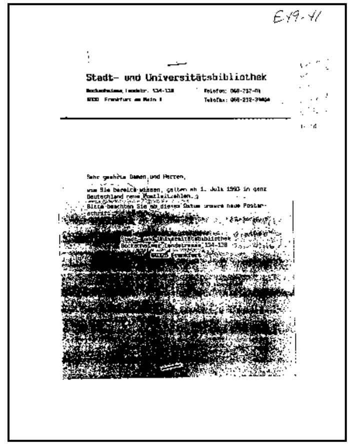






CREATIVE	MARKET	ADVERTISING VEHICLE	DATE
MAIL BOX (GENERIC PRICE)	ARKANSAS	NEWSPAPER	10/10/89
BEACH (GENERIC PRICE)	ARKANSAS	SUNDAY SUPPLEMENT	11/5/89
EAGLE	INDIANA	SUNDAY SUPPLEMENT	11/5/89
EAGLE (NEW GENERIC PRICE)	INDIANA	NEWSPAPER	10/26/89
EAGLE (NEW LOW PRICE)	INDIANA MARKETS	COM	12/89-1/90
RED FACES	ARKANSAS MARKETS	COM	1/90
BEACH	PORTLAND, OREGONIAN	SUNDAY SUPPLEMENT	11/5/89
MAIL BOX	PORTLAND & EUGENE	NEWSPAPER	10/10/89
BEACH	OREGON, EUGENE REGISTER GUARD	SUNDAY SUPPLEMENT	11/15/89
GIRL GUY HEAD-ON	ARKANSAS MARKETS	COM	1/89-12/89
GIRL GUY HEAD-ON REG. & LIGHTS	OREGON MARKETS	SUNDAY SUPPLEMENT	1/89-12/89
"CORVETTE"	PORTLAND, OREGON	SUNDAY SUPPLEMENT	11/26/89
"CORVETTE"	EUGENE, OREGON	SUNDAY SUPPLEMENT	11/26/89
"CORVETTE"	ARKANSAS MARKETS	SUNDAY SUPPLEMENT	11/26/89
"BEACH – LOW PRICE"	ARKANSAS	NEWSPAPER	11/20
"BEACH"	OREGON MARKETS	NEWSPAPER	11/20

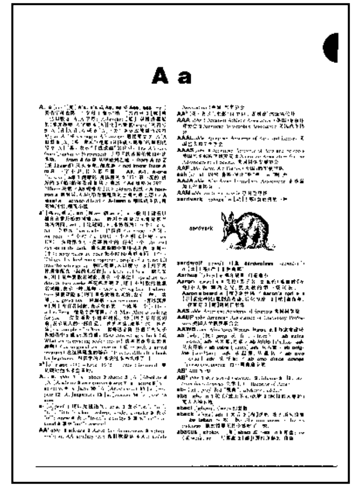
October 18, 1989



Stadt- und Universitätsbibliothek

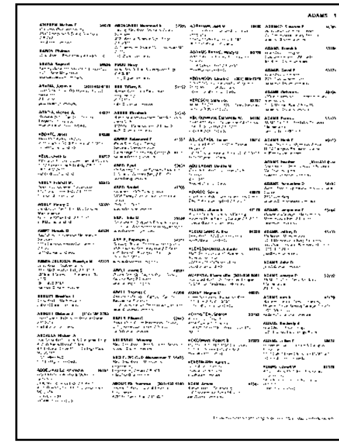
Frankfurt am Main

Sehr geehrte Damen und Herren,



A a

aardvark



The Washington Post

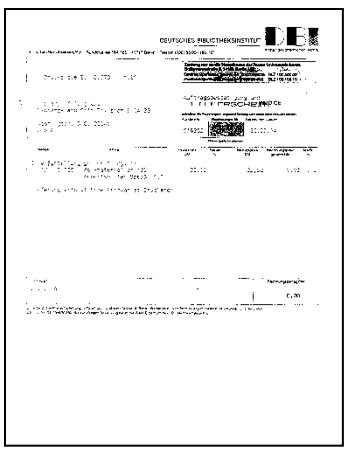
After Zarqawi, No Clear Path In Weary Iraq

The Young Apprentice

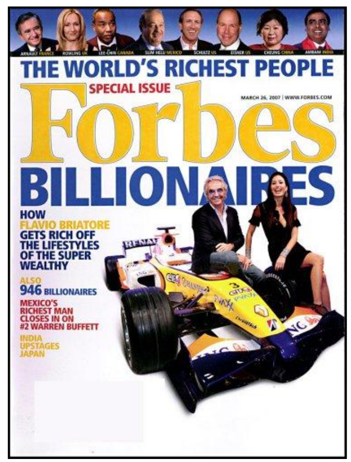
How U.S. Forces Found Iraq's Most-Wanted Man

Kaine Delays Execution Of Inmate for 6 Months

FDA Approves Vaccine That Should Prevent Most Cervical Cancers



DEUTSCHES BIBLIOTHEKSINSTITUT



THE WORLD'S RICHEST PEOPLE

SPECIAL ISSUE

Forbes

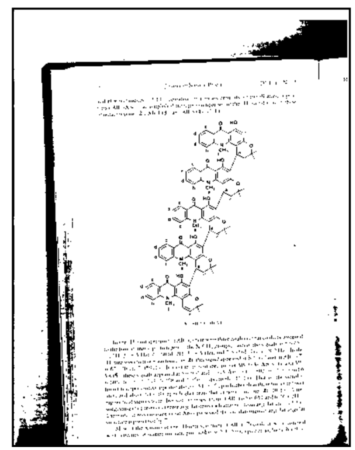
BILLIONAIRES

HOW FLAVIO BRIATORE GETS RICH OFF THE LIFESTYLES OF THE SUPER WEALTHY

ALSO 946 BILLIONAIRES

MEXICO'S RICHEST MAN CLOSES IN ON #2 WARREN BUFFETT

INDIA UPSTAGES JAPAN



FINANCIAL TIMES

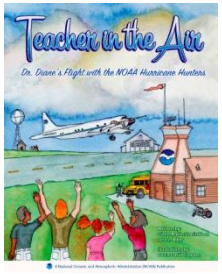
Collateral damage

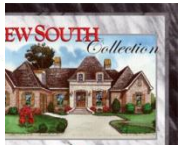
Iraq braced as Saddam rejects Bush ultimatum

Fed holds rates but signals uncertain future

US probes Abold collusion claims

Other “Documents”

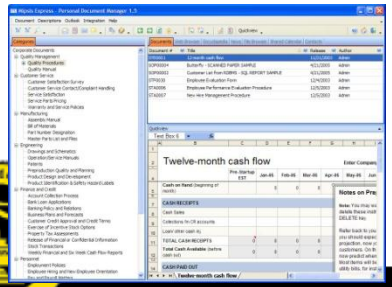


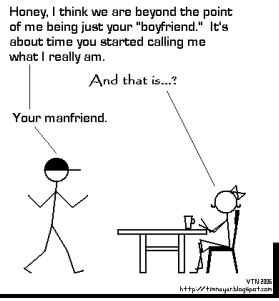


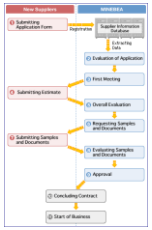
Parliamentary Assembly
Assemblée parlementaire

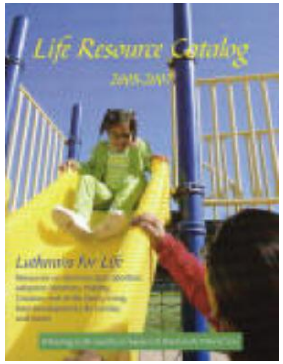
COUNCIL OF EUROPE CONSEIL DE L'EUROPE

I E T F









ที่
ทูซอนริโก
5 กันยายน 2521

ผมคิดถึงเพื่อนรัก

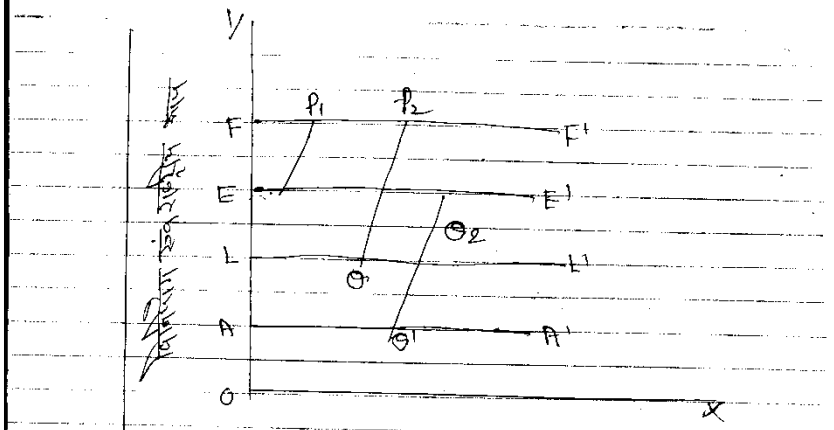
ตั้งแต่กลับจากเมืองไทยไปอยู่อเมริกากับพ่อแม่ได้มาอยู่กับเพื่อนเกือบครบเดือนแล้ว แต่ก็ไม่เคยได้รับจดหมายตอบจากเพื่อนเลย จะเป็นเพราะว่าเพื่อนไม่ได้รับจดหมายของฉันหรือไงก็ไม่ทราบ ฉันคิดถึงเพื่อนมาก แต่ที่ต้องจากกันก็เพื่อเรียนหนังสือ และหวังว่าเพื่อนคงจะตั้งใจเรียนหนังสือเหมือนกัน อย่าเล่นมากจนเกินไป

ฉันหวังใจว่าเพื่อนคงมีความสัมพันธ์ที่ดีกับครอบครัวของเพื่อนเสมอ ถึงแม้ว่าเพื่อนจะไม่ได้อยู่ด้วยกันแล้วก็ตาม ขอให้เพื่อนรักษาความสัมพันธ์ที่ดีนี้ไว้ตลอดไป ถึงแม้ว่าเพื่อนจะมีเรื่องไม่สบายใจก็ขอให้ปรึกษาครอบครัวของเพื่อนเสมอ ฉันขอแสดงความเสียใจด้วยที่เพื่อนต้องเสียคุณพ่อไป ขอให้เพื่อนทำใจดีๆ ไว้ ถึงแม้ว่าคุณพ่อจะไม่อยู่แล้วก็ตาม แต่ท่านก็ยังอยู่ในใจของเพื่อนเสมอ ฉันขอปลอบใจเพื่อนมากๆ ไม่ต้องเสียใจมากนะ ถึงแม้ว่าเมื่อก่อนเพื่อนจะเคยเป็นคนที่ไม่ค่อยมีความอดทนเลย แต่ฉันก็เชื่อว่าเพื่อนจะต้องอดทนได้ ฉันขอให้เพื่อนทำใจดีๆ ไว้ ถึงแม้ว่าคุณพ่อจะไม่อยู่แล้วก็ตาม แต่ท่านก็ยังอยู่ในใจของเพื่อนเสมอ

ฉันอยากเห็นหน้าเพื่อนเต็มทีแล้ว ฉันหวังว่าเพื่อนคงจะมีรูปส่งมาให้ฉันดูบ้าง ฉันจะได้ดีใจ ถึงแม้ว่าเพื่อนจะไม่มีรูปส่งมาให้ฉันดู ฉันก็ดีใจ ขอให้เพื่อนมีความสุขมากๆ นะ ฉันจะรอจดหมายของเพื่อนอยู่ตลอดเวลา ถึงแม้ว่าเพื่อนจะไม่ได้เขียนจดหมายมาหาฉันก็ตาม ฉันก็ยังรอจดหมายของเพื่อนอยู่ตลอดเวลา ขอให้เพื่อนมีความสุขมากๆ นะ ฉันจะรอจดหมายของเพื่อนอยู่ตลอดเวลา ถึงแม้ว่าเพื่อนจะไม่ได้เขียนจดหมายมาหาฉันก็ตาม ฉันก็ยังรอจดหมายของเพื่อนอยู่ตลอดเวลา

ขอให้เพื่อนมีความสุขมากๆ นะ ฉันจะรอจดหมายของเพื่อนอยู่ตลอดเวลา ถึงแม้ว่าเพื่อนจะไม่ได้เขียนจดหมายมาหาฉันก็ตาม ฉันก็ยังรอจดหมายของเพื่อนอยู่ตลอดเวลา

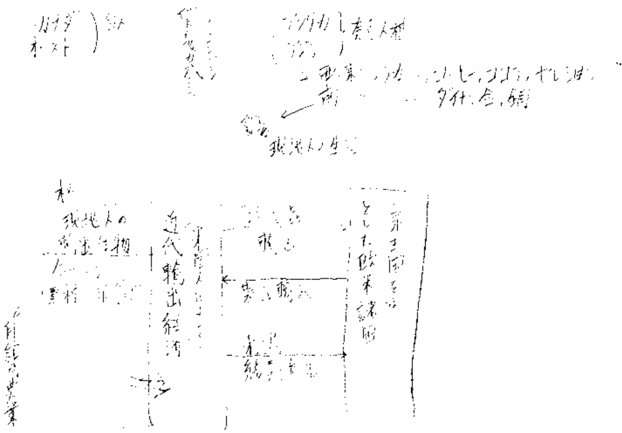
รักและคิดถึงเพื่อนเสมอ



II 途上国の社会経済的特質

1 基本的構造 ～ 従属型モノ経済

イギリスは歪んだ経済構造を作り出した。これは植民地型モノ経済と呼ばれるもので、他の白人居住植民地（オーストラリア、NZ、カナダ、[illegible]）が自分の発展により急速に[illegible]のに[illegible]植民地は停滞を余儀なくされた。



植民地型モノ経済[illegible]特徴[illegible]
[illegible]

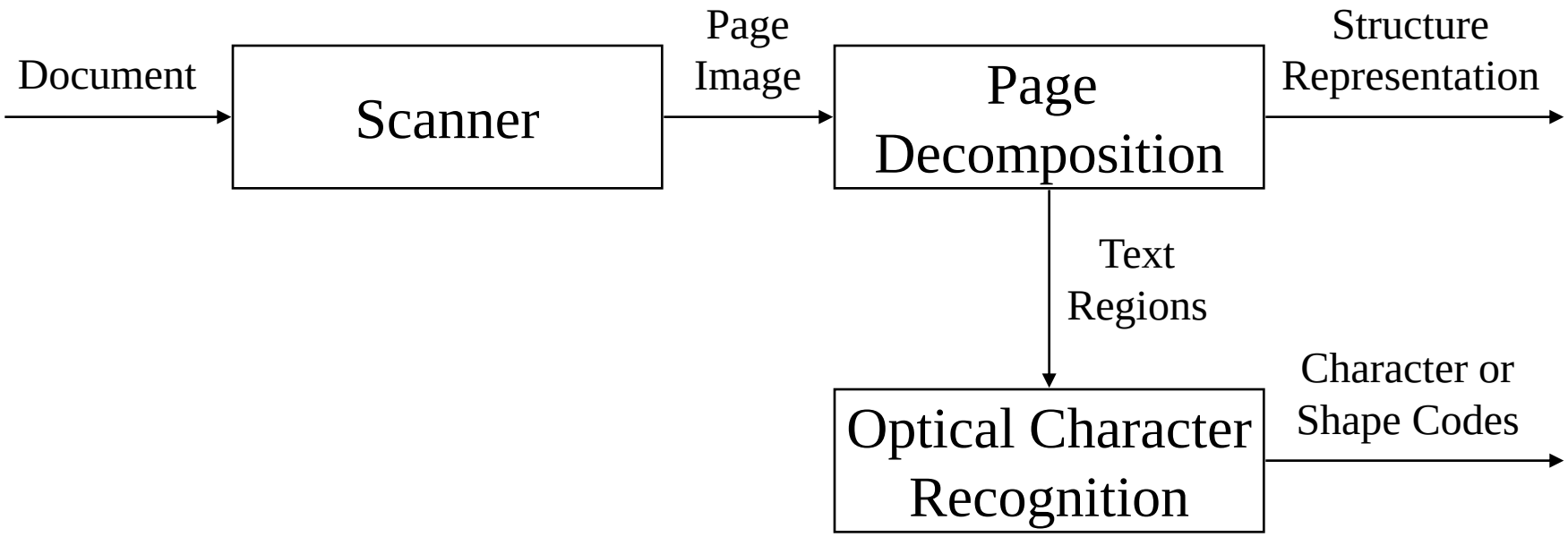
6. The Three Kingdoms and the Six Dynasty

p85 férment	n.	불안, 동요, 정치적 동요
usúrp	v.	강탈하다
incéssant	a.	끊임없는
shrink	v.	움츠러들다, 줄어들다, 기가 죽다 (p.p. shrunken)
rashly	ad.	무모하게, 경솔하게, 성급하게
mássacre	v.	학살하다
flee	v.	달아나다, 도피하다 (p. fled)
tribe	n.	부족, 종족
abándon	v.	버리다, 포기하다
nomádic	a.	유목민의, 유목 생활을 하는
* Sínicize	v.	중국화하다, 중국식으로 만들다
cávalry	n.	기병대, 기마부대
refugée	n.	망명자, 피난자, 난민
perpétual	a.	영구한, 부단한
tórmoil	n.	혼란, 동요, 불안
p87 undermind	v.	약화시키다, 서서히 퇴화시키다
* mónastery	n.	수도원, 사원, 사찰(절)
vast	a.	광대한, 거대한, 엄청난
* propórtion	n.	크기, 비례, 비율
realm /relm/	n.	왕국, 영역
bureáucracy	n.	관료제, 관료정치, 관료주의
* exért	v.	(힘·권력 등을) 행사하다, (영향을) 미치다

〈 Taoism 〉 도교

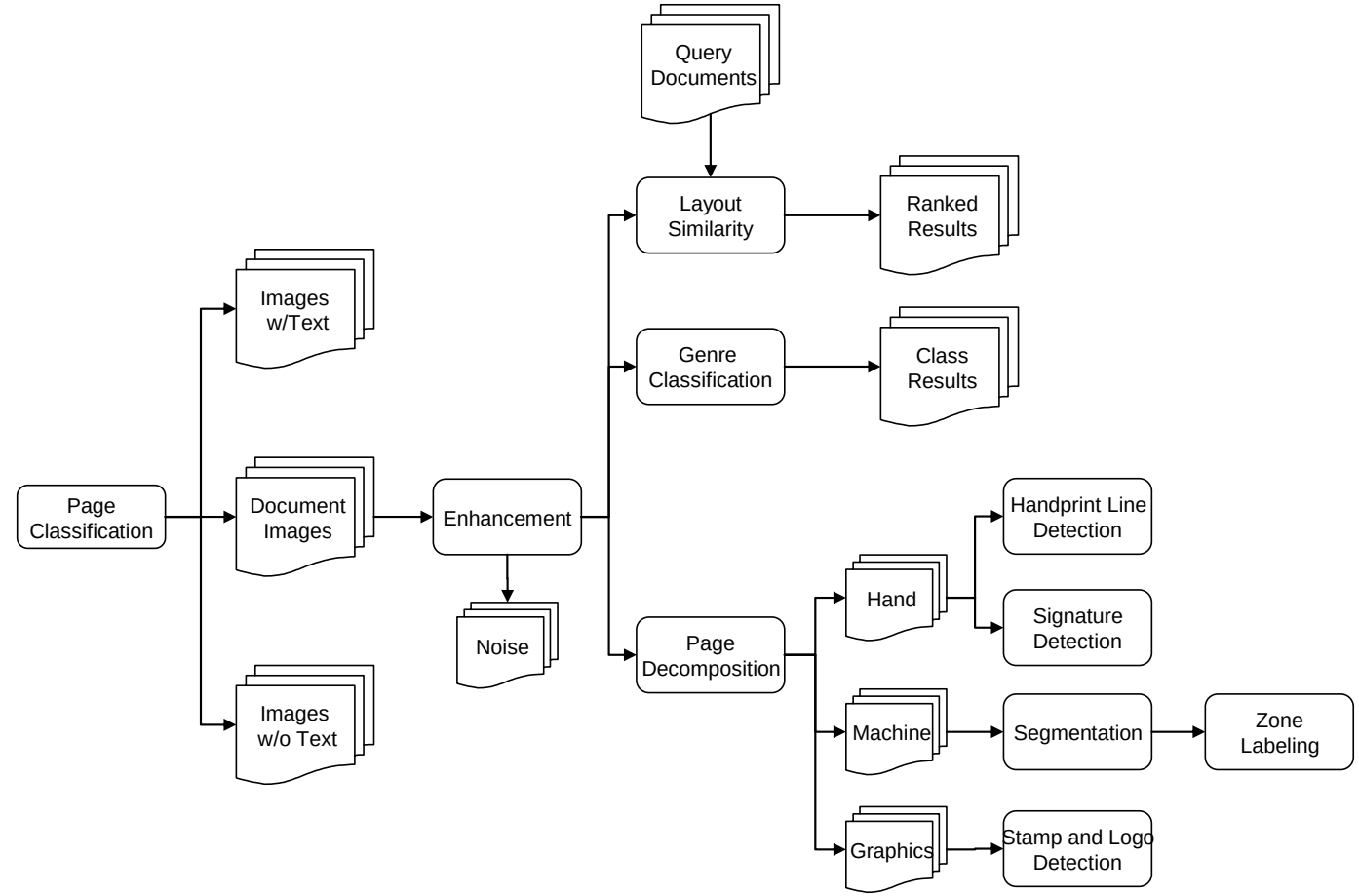
Táoism	n.	도교
*callígraphy	n.	서예, 서도, 달필, 필적
conglomerátion	n.	응집, 집합

Indexing Page Images



Document Image Analysis

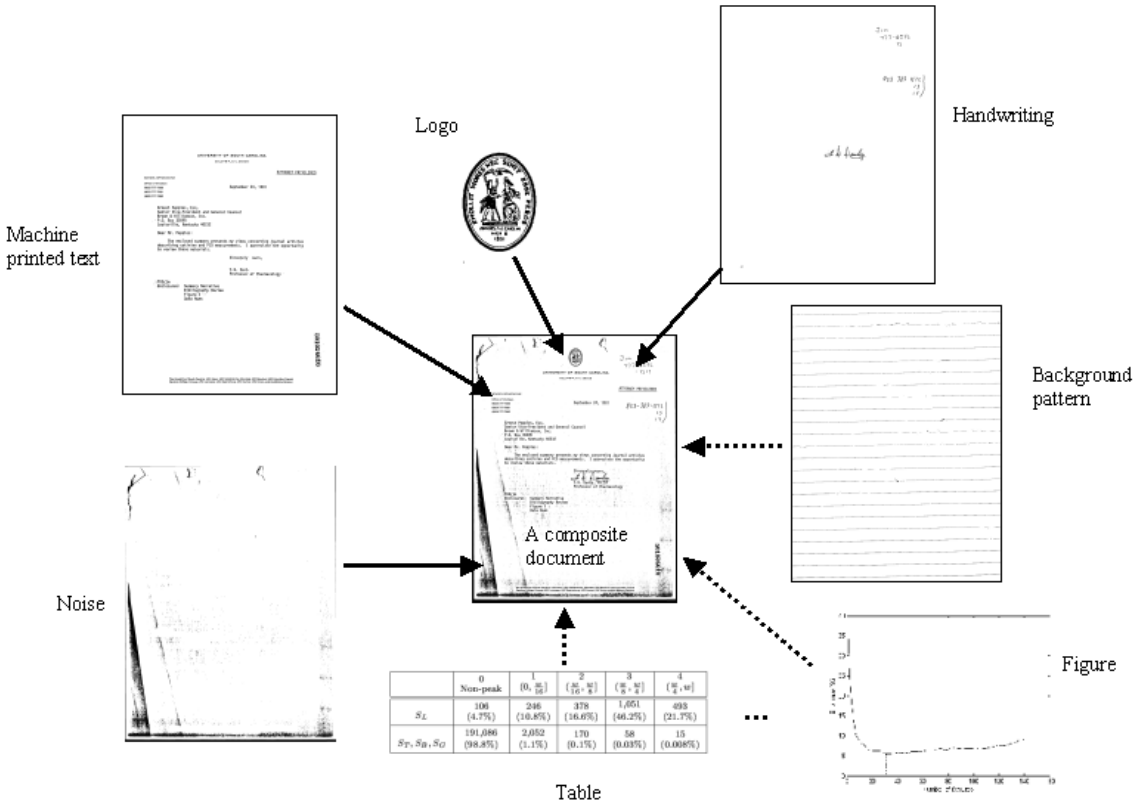
- General Flow:
  - Obtain Image - Digitize
  - Preprocessing
  - Feature Extraction
  - Classification
- General Tasks
  - Logical and Physical Page Structure Analysis
  - Zone Classification
  - Language ID
  - Zone Specific Processing
    - Recognition
    - Vectorization



Target Processing Speed in Seconds

Page Layer Segmentation

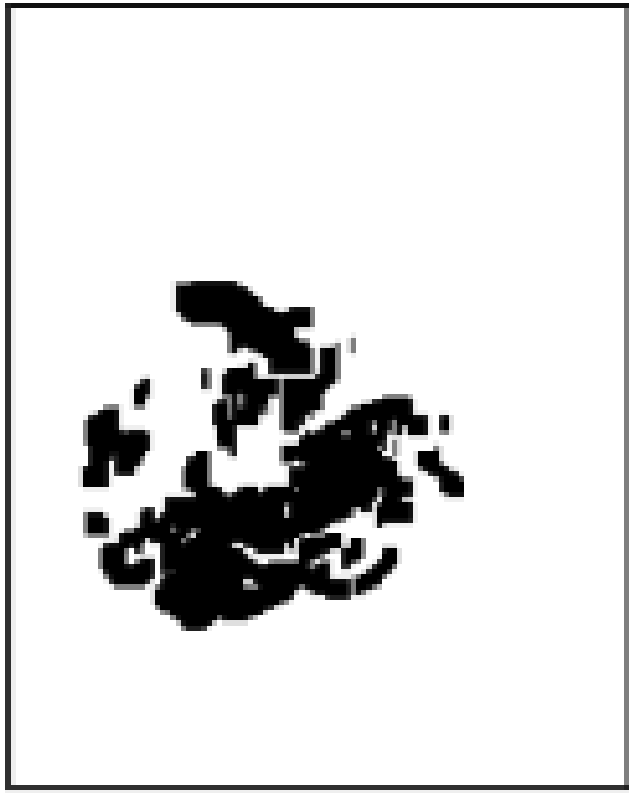
- Document image generation model
  - A document consists many layers, such as handwriting, machine printed text, background patterns, tables, figures, noise, etc.

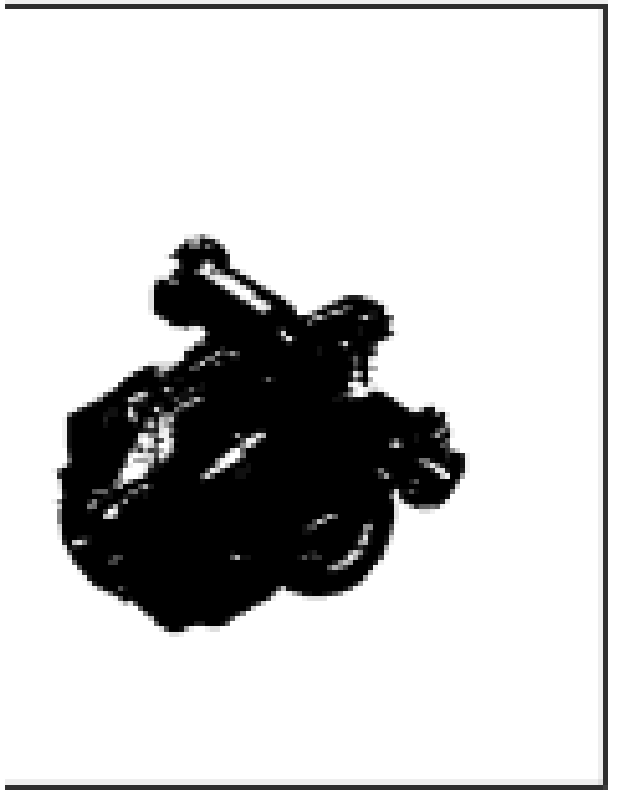


Page Analysis

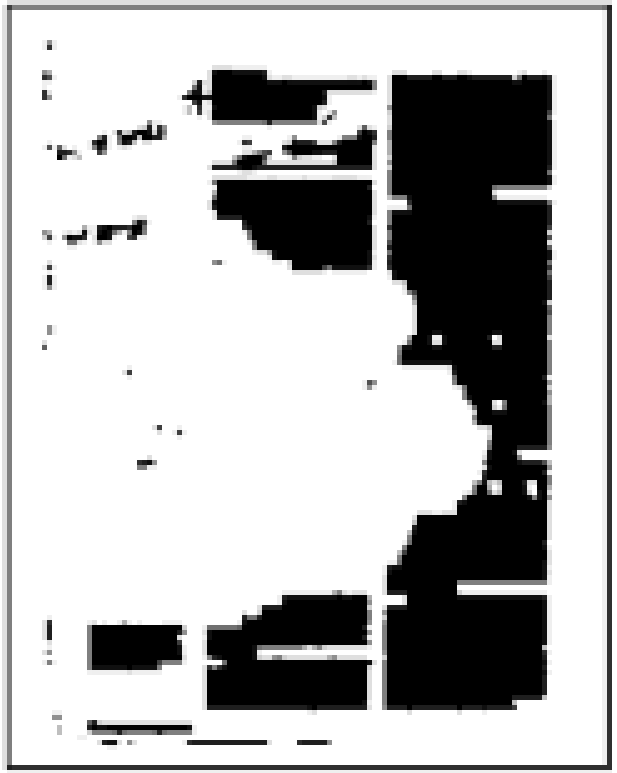
- Skew correction
  - Based on finding the primary orientation of lines
- Image and text region detection
  - Based on texture and dominant orientation
- Structural classification
  - Infer logical structure from physical layout
- Text region classification
  - Title, author, letterhead, signature block, etc.

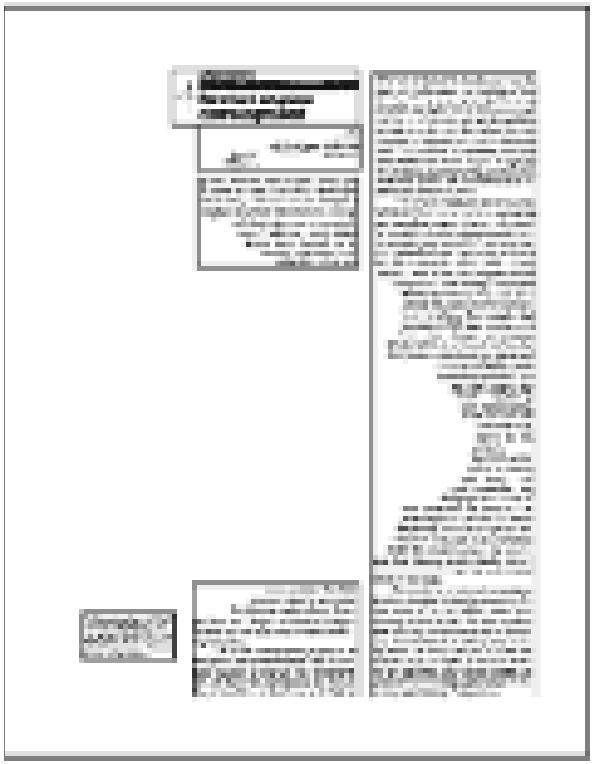
Image Detection





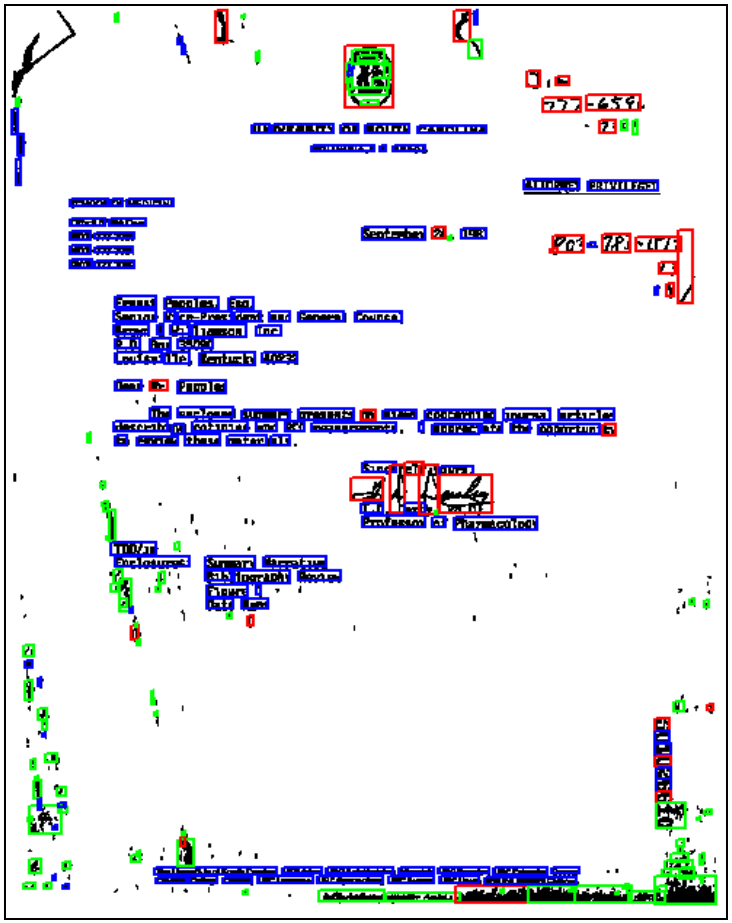
Text Region Detection



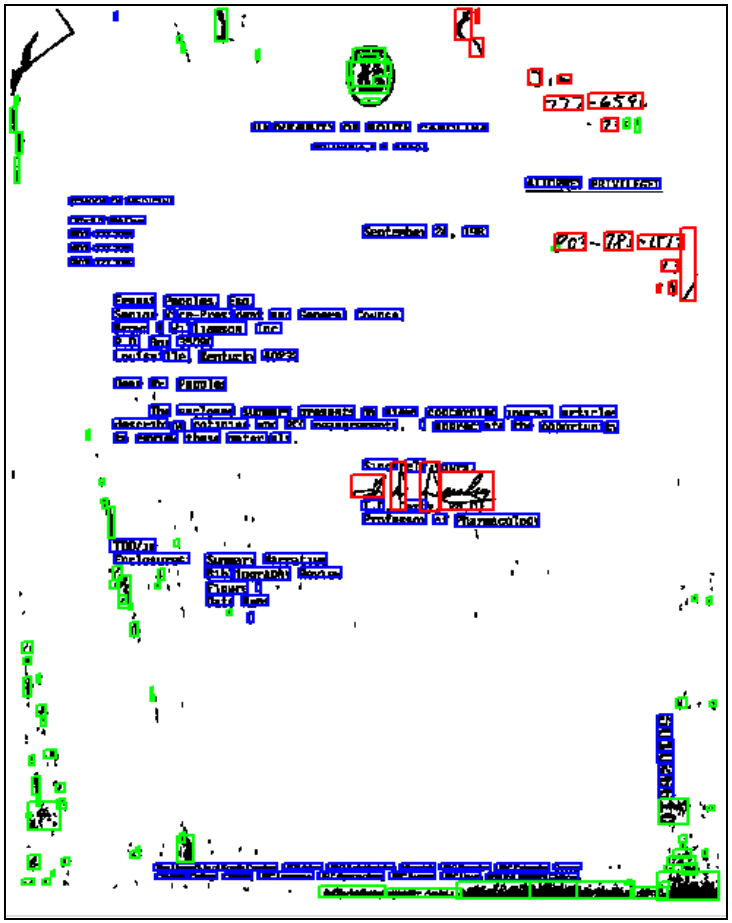


More Complex Example

Printed text
Handwriting
Noise

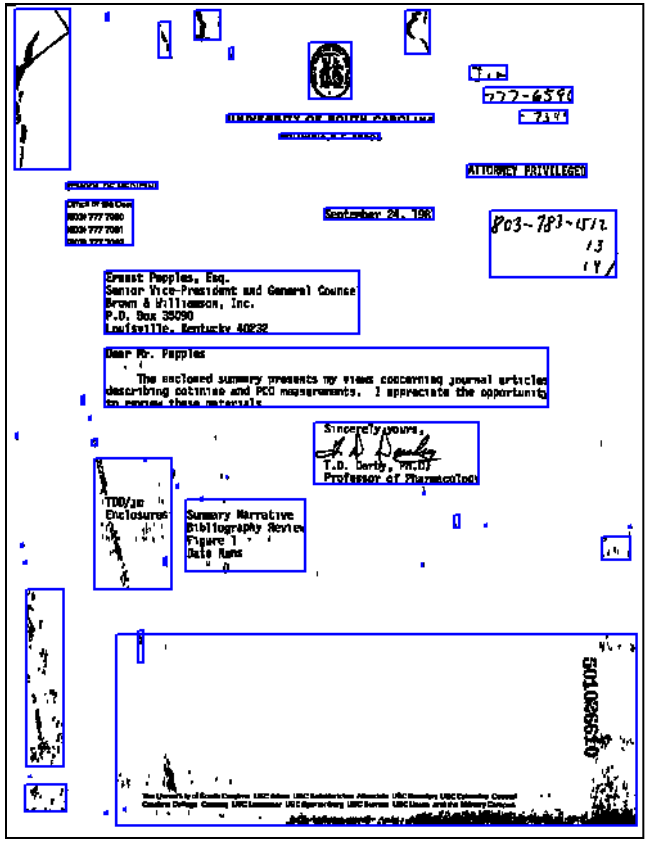


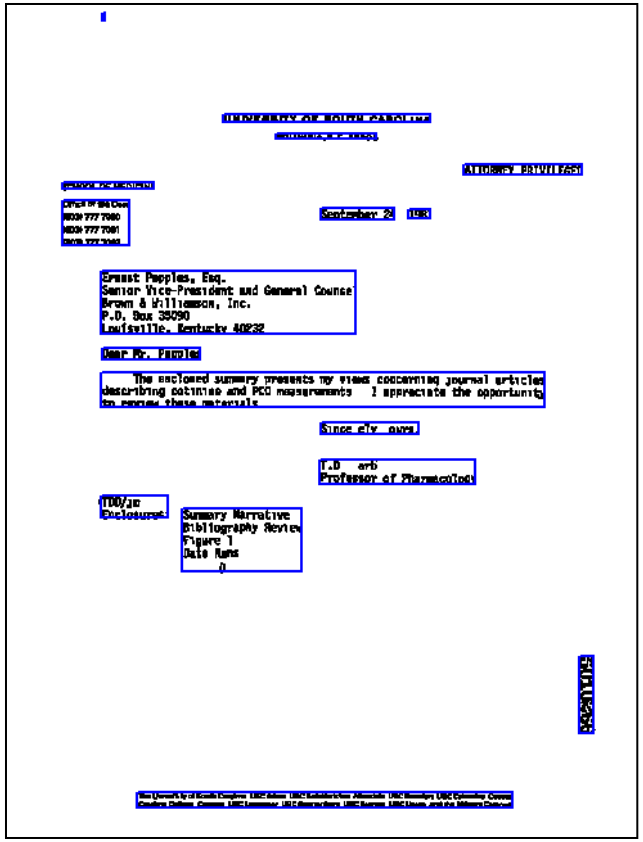
Before MRF-based postprocessing



After MRF-based postprocessing

Application to Page Segmentation





Before enhancement

After enhancement

Language Identification

- Language-independent skew detection
  - Accommodate horizontal and vertical writing
- Script class recognition
  - Asian scripts have blocky characters
  - Connected scripts can’t be segmented easily
- Language identification
  - Shape statistics work well for western languages
  - Competing classifiers work for Asian languages

What about handwriting?

Optical Character Recognition

- Pattern-matching approach
  - Standard approach in commercial systems
  - Segment individual characters
  - Recognize using a neural network classifier
- Hidden Markov model approach
  - Experimental approach developed at BBN
  - Segment into sub-character slices
  - Limited lookahead to find best character choice
  - Useful for connected scripts (e.g., Arabic)

OCR Accuracy Problems

- Character segmentation errors
  - In English, segmentation often changes “m” to “rn”
- Character confusion
  - Characters with similar shapes often confounded
- OCR on copies is much worse than on originals
  - Pixel bloom, character splitting, binding bend
- Uncommon fonts can cause problems
  - If not used to train a neural network

Improving OCR Accuracy

- Image preprocessing
  - Mathematical morphology for bloom and splitting
  - Particularly important for degraded images
- “Voting” between several OCR engines helps
  - Individual systems depend on specific training data
- Linguistic analysis can correct some errors
  - Use confusion statistics, word lists, syntax, …
  - But more harmful errors might be introduced

OCR Speed

- Neural networks take about 10 seconds a page
  - Hidden Markov models are slower
- Voting can improve accuracy
  - But at a substantial speed penalty
- Easy to speed things up with several machines
  - For example, by batch processing - using desktop computers at night

Problem: Logical Page Analysis (Reading Order)

- Can be hard to guess in some cases
  - Newspaper columns, figure captions, appendices, …
- Sometimes there are explicit guides
  - “Continued on page 4” (but page 4 may be big!)
- Structural cues can help
  - Column 1 might continue to column 2
- Content analysis is also useful
  - Word co-occurrence statistics, syntax analysis

Information Retrieval on OCR

- Requires robust ways of indexing
- Statistical methods with large documents work best
- Key Evaluations
  - Success for high quality OCR (Croft et al 1994, Taghva 1994)
  - Limited success for poor quality OCR (1996 TREC, UNLV)

N-Grams

- Powerful, Inexpensive statistical method for characterizing populations
- Approach
  - Split up document into n-character pairs fails
  - Use traditional indexing representations to perform analysis
  - “DOCUMENT” -> DOC, OCU, CUM, UME, MEN, ENT
- Advantages
  - Statistically robust to small numbers of errors
  - Rapid indexing and retrieval
  - Works from 70%-85% character accuracy where traditional IR fails

Matching with OCR Errors

- Above 80% character accuracy, use words
  - With linguistic correction
- Between 75% and 80%, use n-grams
  - With n somewhat shorter than usual
  - And perhaps with character confusion statistics
- Below 75%, use word-length shape codes

Handwriting Recognition

- With stroke information, can be automated
  - Basis for input pads
- Simple things can be read without strokes
  - Postal addresses, filled-in forms
- Free text requires human interpretation
  - But repeated recognition is then possible

Additional Reading

- A. Balasubramanian, et al. Retrieval from Document Image Collections, *Document Analysis Systems VII*, pages 1-12, 2006.

- D. Doermann. The Indexing and Retrieval of Document Images: A Survey. *Computer Vision and Image Understanding*, 70(3), pages 287-298, 1998.

Some Applications

- Legacy Tobacco Documents Library
  - http://legacy.library.ucsf.edu/
- Google Books
  - http://books.google.com/
- George Washington's Papers
  - http://ciir.cs.umass.edu/irdemo/hw-demo/

Agenda

- Beyond Text, but still language
  - Scanned documents
  - ➢Speech
- Beyond Text, but still information
  - Images
  - Video
- Beyond text to data

Encoding Audio

- Nyquist theorem: sampling rate
- Channel characteristics
  - Bandpass
  - Amplitude resolution
- Compression
  - Lossless
  - Perceptually-based lossy (e.g., .mp3)

Description Strategies

- Transcription
  - Manual transcription (with optional post-editing)
- Annotation
  - Manually assign descriptors to points in a recording
  - Recommender systems (ratings, link analysis, …)
- Associated materials
  - Interviewer's notes, speech scripts, producer's logs
- Automatic
  - Create access points with automatic speech processing

Detectable Speech Features

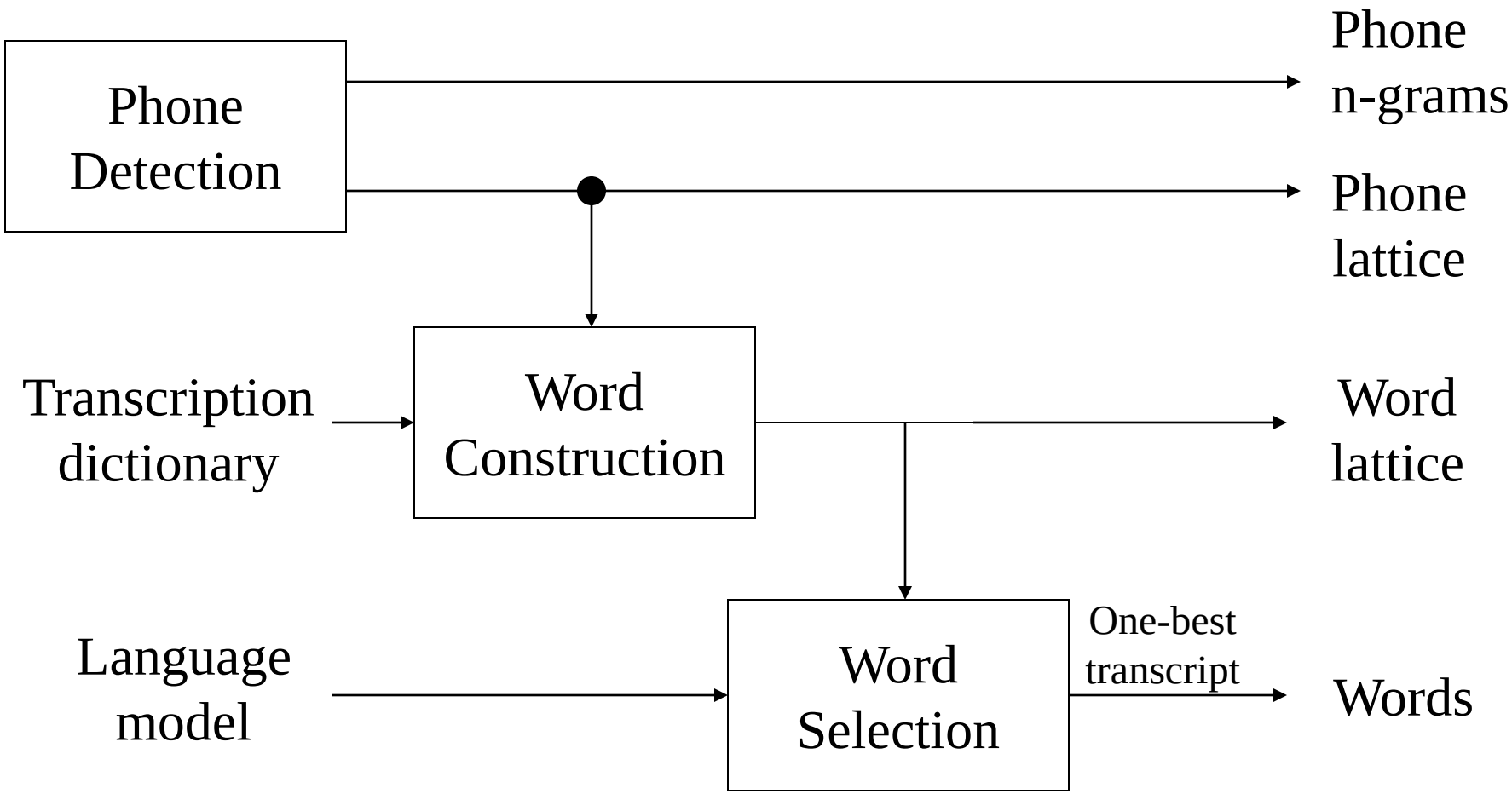
- Content
  - Phonemes, one-best word recognition, n-best
- Identity
  - Speaker identification, speaker segmentation
- Language
  - Language, dialect, accent
- Other measurable parameters
  - Time, duration, channel, environment

How Speech Recognition Works

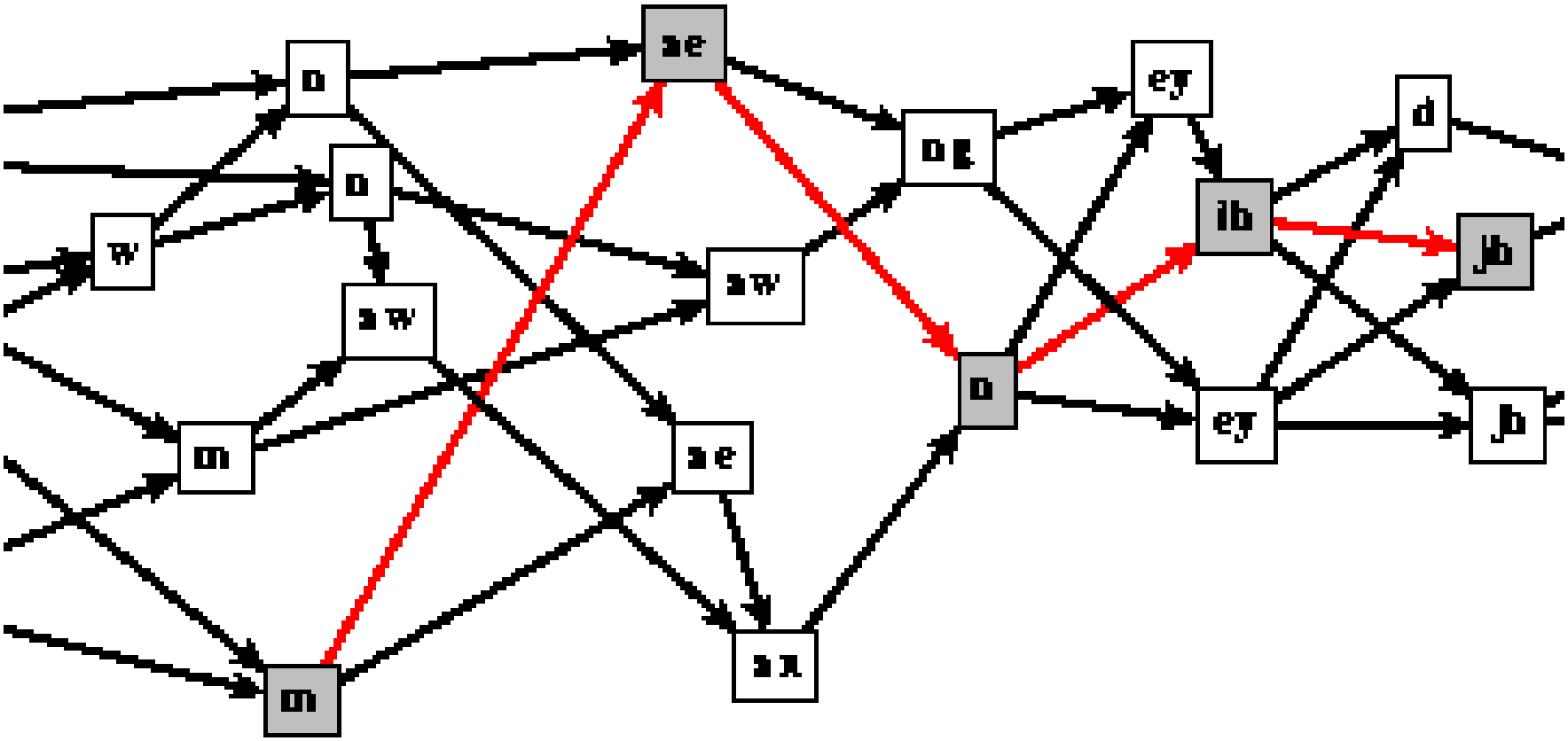
- Three stages
  - What sounds were made?
    - Convert from waveform to subword units (phonemes)
  - How could the sounds be grouped into words?
    - Identify the most probable word segmentation points
  - Which of the possible words were spoken?
    - Based on likelihood of possible multiword sequences

- These can be learned from “training data”
  - Hill-climbing sequence models
    - (a “Hidden Markov Model”)

Using Speech Recognition



Phone Lattice



Phoneme Trigrams

- Manage -> m ae n ih jh
  - Dictionaries provide accurate transcriptions
    - But valid only for a single accent and dialect
  - Rule-base transcription handles unknown words
- Index every overlapping 3-phoneme sequence
  - m ae n
  - ae n ih
  - n ih jh

Speech Retrieval Evaluations

- 1996-1998: TREC SDR
  - EN broadcast news / EN queries
- 1997-2004: TDT
  - EN+CN+AR broadcast news / Query by example
- 2003-2004: CLEF CL-SDR
  - EN broadcast news / Many query languages
- 2005-2007: CLEF CL-SR
  - EN+CZ interviews / Many query languages

Key Results from TREC/TDT

- Recognition and retrieval can be decomposed
  - Word recognition/retrieval works well in English
- Retrieval is robust with recognition errors
  - Up to 40% word error rate is tolerable
- Retrieval is robust with segmentation errors
  - Vocabulary shift/pauses provide strong cues

A Richer View of Speech

- Speaker identification
  - Known speaker and “more like this” searches
  - Gender detection for search and browsing
- Topic segmentation
  - Vocabulary shift, cue words
  - More natural breakpoints for browsing
- Speaker segmentation
  - Visualize turn-taking behavior for browsing
  - Classify turn-taking patterns for searching

Speaker Identification

- Gender
  - Classify speakers as male or female
- Identity
  - Detect speech samples from same speaker
  - To assign a name, need a known training sample
- Speaker segmentation
  - Identify speaker changes
  - Count number of speakers

Competing Demands on the Interface

- Query must result in a manageable set
  - But users prefer simple query interfaces
- Selection interface must show several segments
  - Representations must be compact, but informative
- Rapid examination should be possible
  - But complete access to the recordings is desirable

BBN Radio News Retrieval

AT&T Radio News Retrieval

QUERY: What is the status of the trade deficit with Japan? SEARCH CLEAR

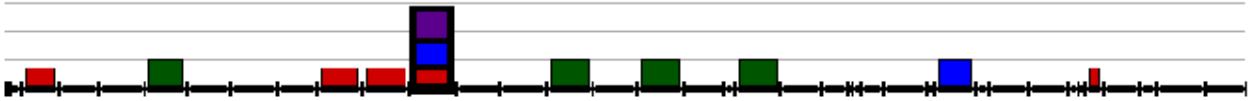
RESULTS – "What is the status of the trade deficit with Japan"

RANK	PROGRAM	DATE	STORY	SCORE	LENGTH	HITS
1	NPR All Things Considered	05/31	3	15.63	27.65	6
2	NPR All Things Considered	05/10	15	13.89	512.12	16
3	NPR/PRI Marketplace	06/11	1	13.82	166.10	11
4	ABC World News Now	06/13	6	13.11	30.00	3
5	NPR All Things Considered	05/21	1	11.11	13.62	3
6	NPR All Things Considered	05/31	3	10.92	17.02	3
7	NPR/PRI Marketplace	06/11	3	10.87	30.00	1
8	CNN Headline News	06/07	18	9.83	183.55	6
9	NPR/PRI Marketplace	06/11	23	9.82	203.21	11
10	NPR/PRI Marketplace	06/14	6	9.41	90.33	4

Prev Doc Next Doc

OVERVIEW – NPR All Things Considered 05/10

deficit
status
japan
trade



ASR TRANSCRIPTS – NPR All Things Considered 05/10

"expanding defense cooperation span is a part of our pacific democracy defense program will strengthen are lines and serve on mutual interest that while president clinton is earth credit for renewing inspecting those ties on his recent trip the administration's amateurs and in a factory posturing on trade disputes"

"buster and those ties and assess state of the president's recent attempt of damage control in nineteen ninety four that lead administration for both a trade war and lost and then declared victory even though present but received nothing the clinton a station shows funk war dead and then contradictory tactics"

"did not work for the force camp and saving deregulation competition and economic reform the result has been an increase in both the bilateral trade deficit and japanese trade nationalism the merchandise trade that has no sacred is anthony here no but i do not agree with president clinton's decision"

"the normal eyes relations with vietnam until they could could have and should receive more returned from vietnam the decision has been made the case is not closed there are many outstanding issues in our relationship with vietnam was shared economic and other enters can only be realized"

"after the outcome achieved fullest possible accounting for a missing servicemen and vietnam must understand that further progress on the field of the a. m. i. a. issue remain are biased bilateral priority now it is simply that i think we all saw to be very forthright flat out but i have fun"

"that out neo from about are commercial relations with china was incredible is right the nineteen ninety four when a funny decided extension of most favored nation status was the best way to promote are long term interest in china

Selection Length: 19.1699 seconds Stop Audio

SpeechBot

Simple Search | **Power Search** | **Help** | FAQ | About SpeechBot | Feedback

Search for: organic food [Search]

Topics: All Topics Dates: All dates

Tip: *Try searching a particular topic instead of "**All Topics**"*

Search Result: **200 matches** for your query Sort results by: Relevance

	Website	Date	Extract from Transcript (*Transcripts based on speech recognition are not exact*)
PLAY extract	The Diane Rehm Show	Mar 14, 2000	...new standards for growing and processing **organic food** the proposal incorporates recommendations that consumer groups and **organic** farmers... Show me more
PLAY extract	PBS Online NewsHour	Dec 21, 2000	...advantage of the exploding demand for **organic** products glickman also said the standards would make things a lot clearer for consumers 1 northern virginia **organic food** shopper... Show me more
PLAY	Public Interest	Aug 28, 2000	...get us on the right direction and but you move for a new now we can support farmers markets weekend the point

20 extracts from **The Diane Rehm Show - Mar 14, 2000** match your search:

organic food

You are here
extract 1 of 20

NEXT EXTRACT

PLAY extract

...waga in new and marching tune and diane ream you as to prevent it the agriculture is proposed new standards for growing and processing **organic food** the proposal incorporates recommendations that consumer groups and **organic** farmers but some say that standard to comply with popular opinion rather than scientific research joining me to discuss **organic food** standards can claim they're against chief marketing standards in and the state years for the u. s. department of agriculture tell a d. on stand and ...

Display 30 seconds of transcript

Extracts from this transcript in order of relevance:

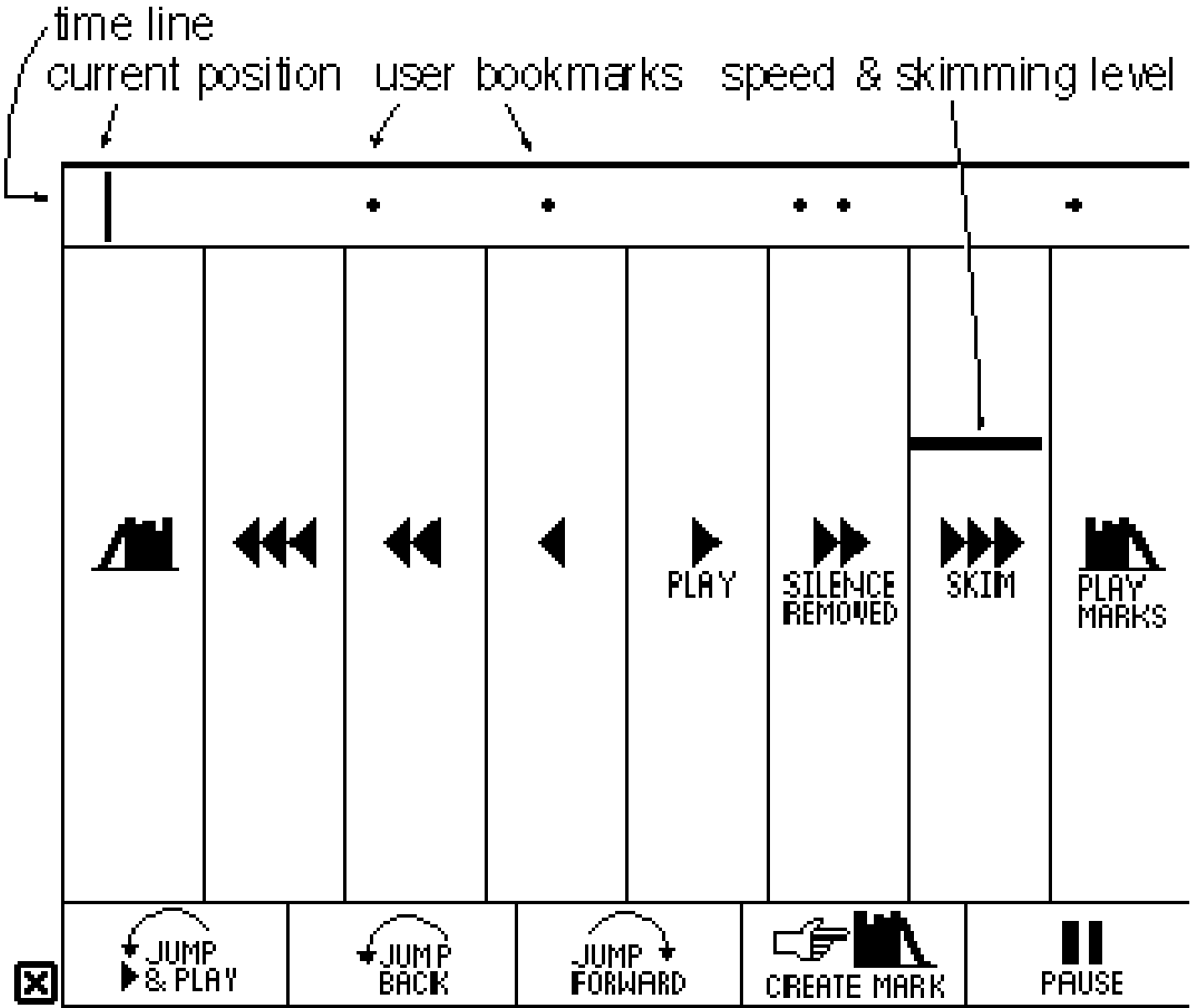
standards for growing and processing organic food the proposal incorporates

PLAY 51 min

The Diane Rehm Show - Mar 14, 2000

Visit this website
Search all transcripts from this website

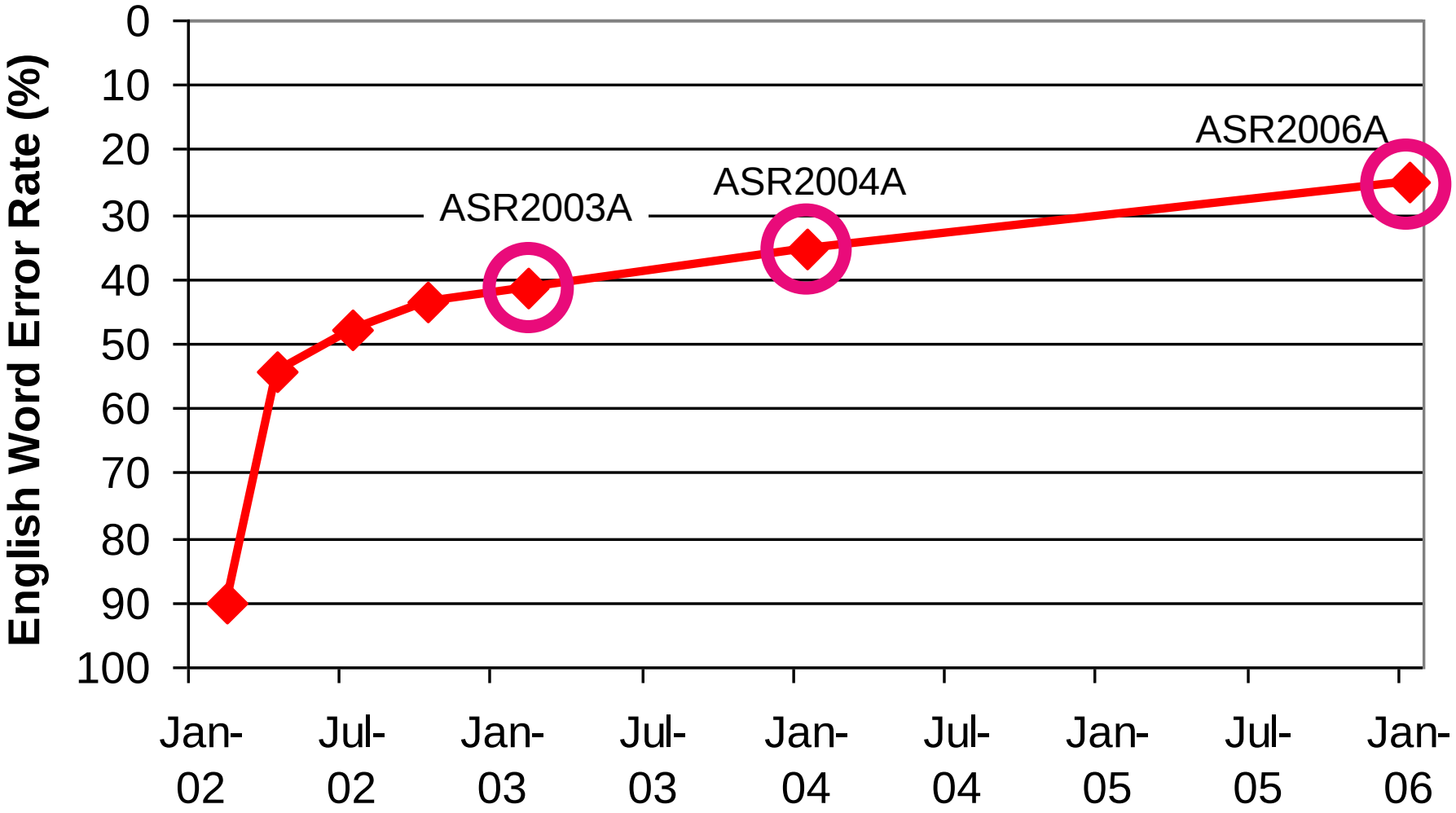
MIT “Speech Skimmer”



Comparison with Text Retrieval

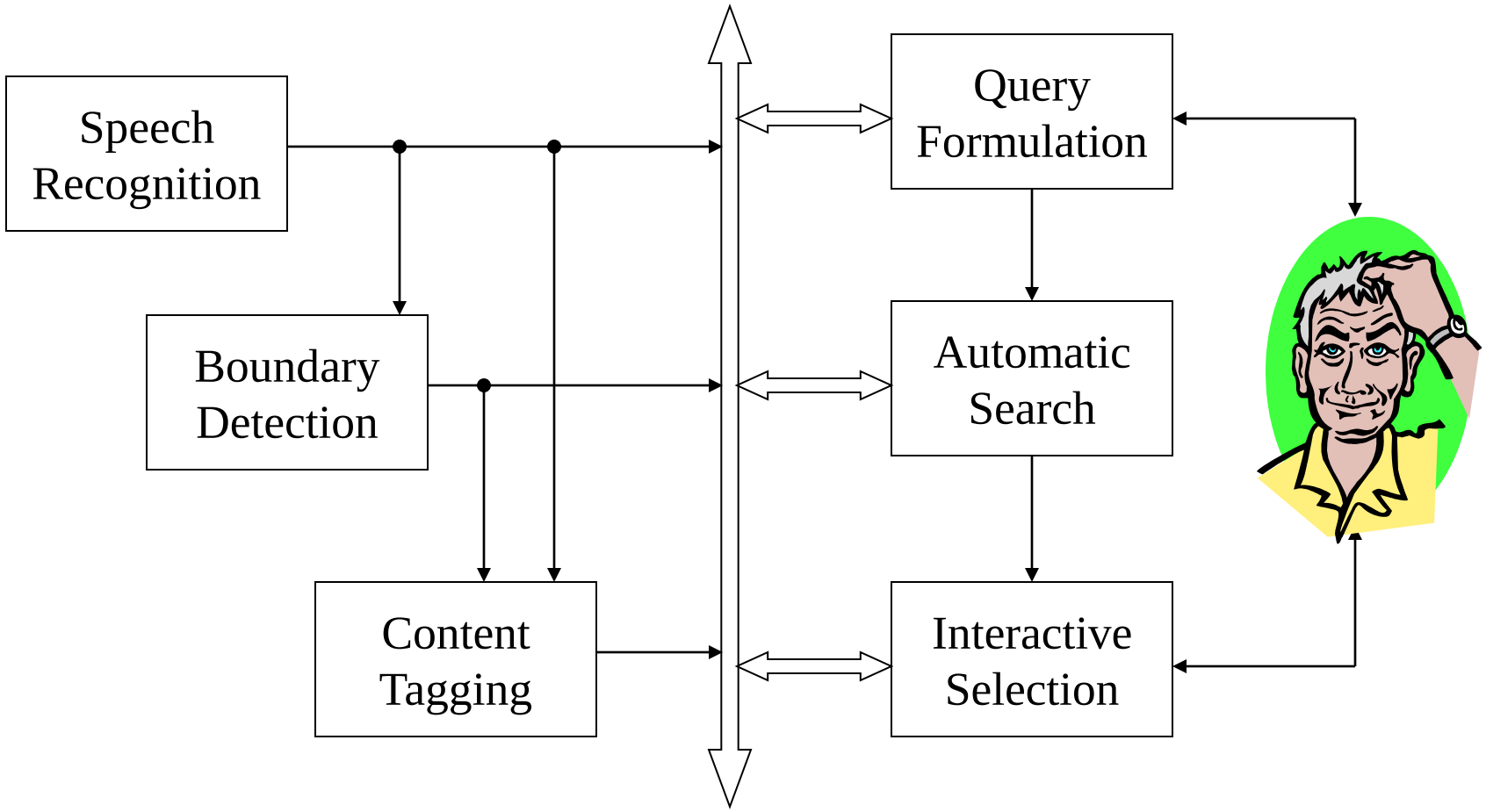
- Detection is harder
  - Speech recognition errors
- Selection is harder
  - Date and time are not very informative
- Examination is harder
  - Linear medium is hard to browse
  - Arbitrary segments produce unnatural breaks

English Transcription Accuracy

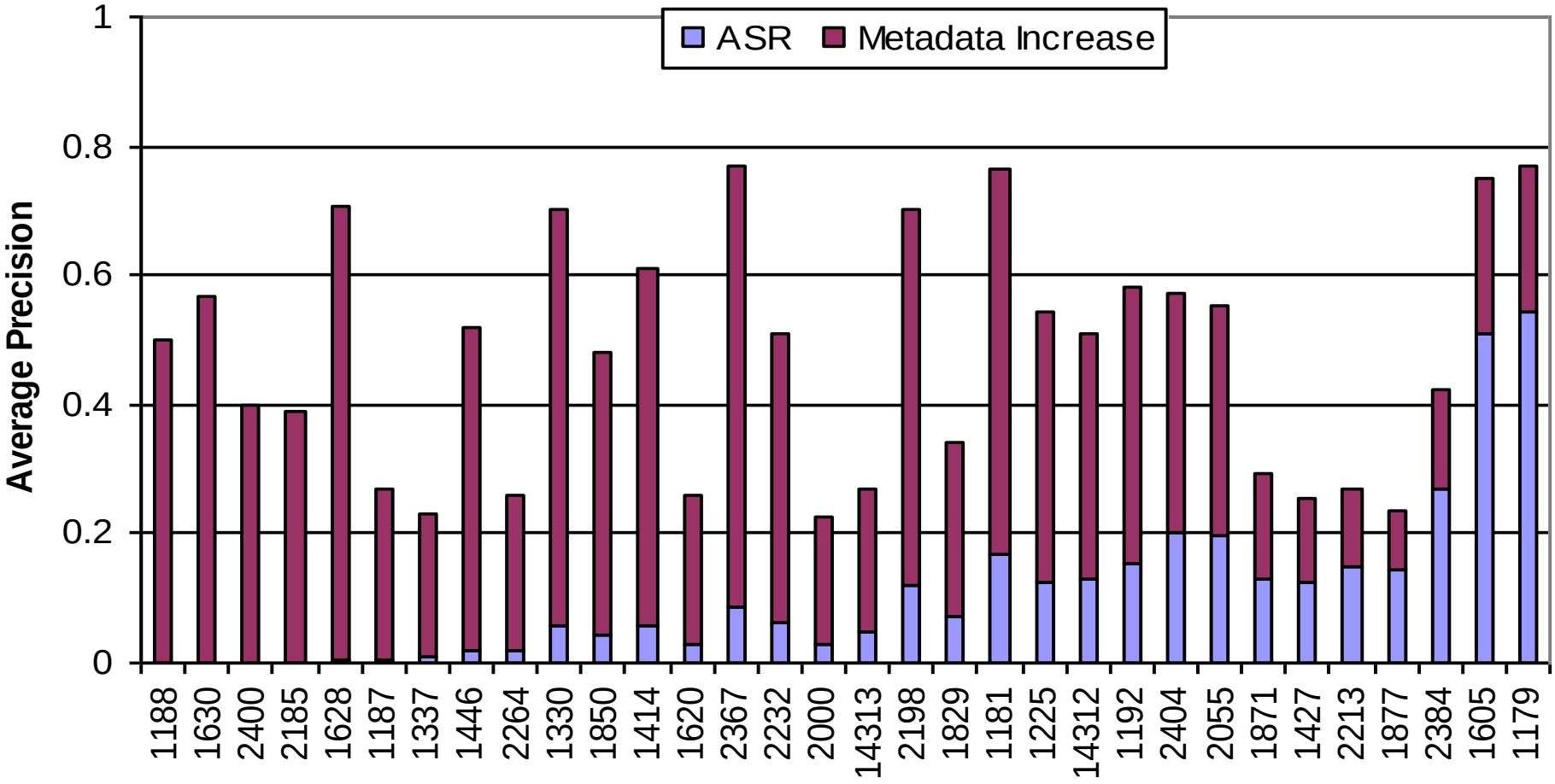


Training: 200 hours from 800 speakers

English Test Collection Design



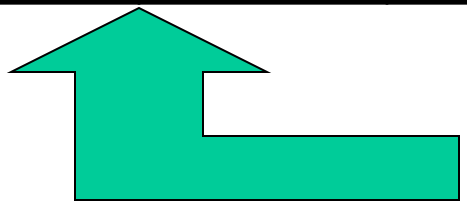
Comparing ASR with Metadata (2005)

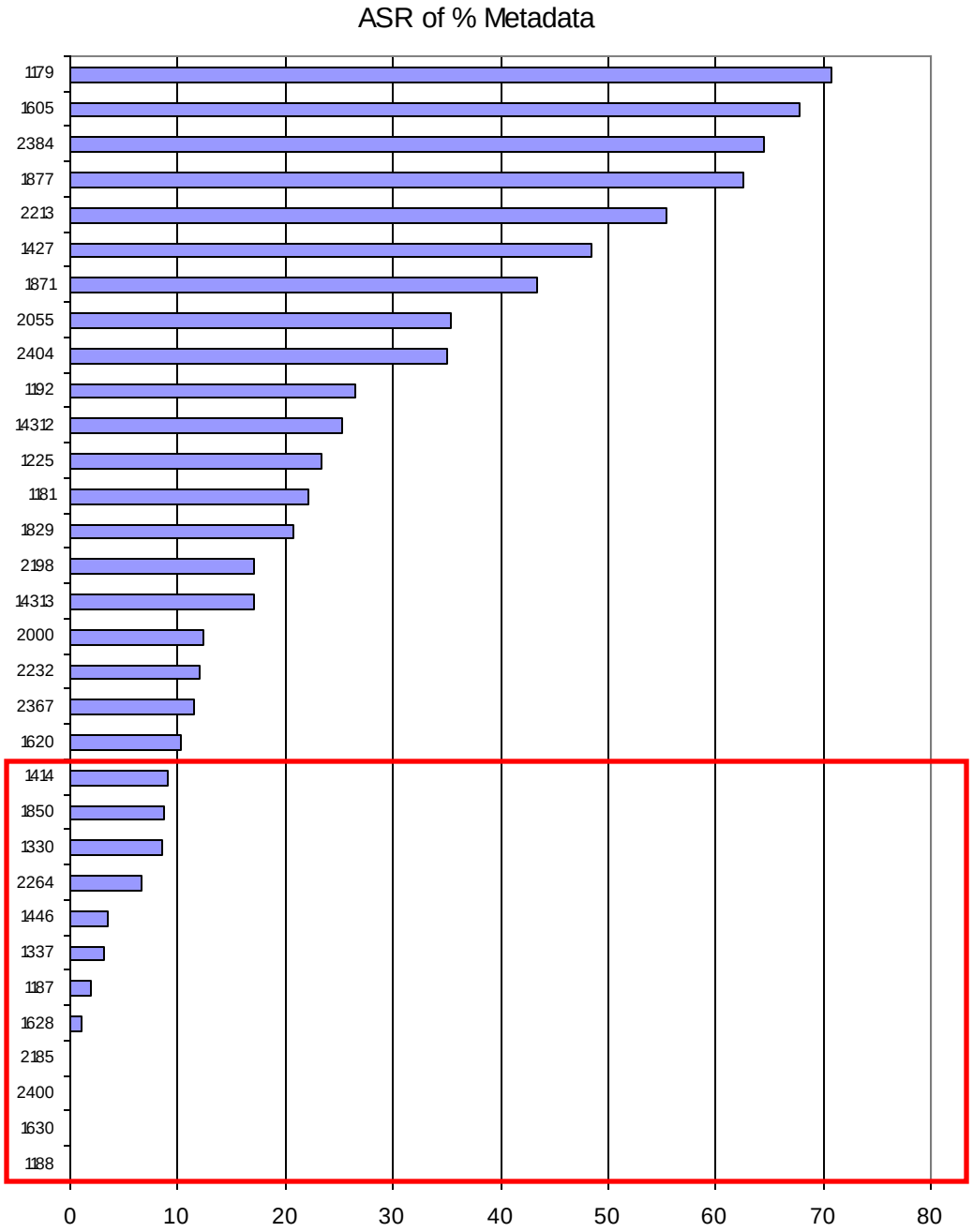


CLEF-2005 training + test – (metadata < 0.2), ASR2004A only, **Title queries**, Inquery 3.1p1

Error Analysis (2005)

Somewhere in ASR (ASR/Metadata)	Only in Metadata	
wallenberg (3/36)* rescue jews		
wallenberg (3/36)	eichmann	
abusive female (8/81) personnel		
minsko (21/71) ghetto underground		
art auschwitz		
labor camps	ig farben	
slave labor	telefunken	aeg
holocaust	sinti roma	
sobibor (5/13) death camp		
witness	eichmann	
jews	volkswagen	





CLEF-2005 training + test – (metadata < 0.2), ASR2004A only, Title queries, Inquery 3.1p1

For More Information

- CLEF Cross-Language Speech Retrieval track
  - http://clef-clsr.umiacs.umd.edu/
- The MALACH project
  - http://malach.umiacs.umd.edu/
- NSF/DELOS Spoken Word Access Group
  - http://www.dcs.shef.ac.uk/spandh/projects/swag

Agenda

- Beyond Text, but still language
  - Scanned documents
  - Speech
- Beyond Text, but still information
  - Images
  - Video
- Beyond text to data

Yahoo! Image Surfer

Computers and Internet:Internet:World Wide We b:Searching the Web

The following is a list of **Image Surfer** categories at **Yahoo**. Stay tuned because more interesting categories are added every week for your viewing pleasure.

Search

Image Surfer Categories

- **Arts**
  Dance, Landscapes, Photography, ...
- **Entertainment**
  Comics and Animation, Rock and Pop, ...
- **People**
  Actors and Actresses, Models, ...
- **Recreation**
  Computer Games, Sports, ...
- **Science**
  Animals, Space and Astronomy, Museums, ...
- **Vehicles**
  Automobiles, Planes, Motorcycles, ...

[Visual Search]
[Image Info]

Flattop, Portuguese AF A7D

[Visual Search]
[Image Info]

Phil's Aircraft Page- Early Flight, Bleriot Monopl...

[Visual Search]
[Image Info]

Diamond Aviation Home Page,

[Visual Search]
[Image Info]

,

[Visual Search]
[Image Info]

Evan's Airliner Photo Index Edition #8, Federal Ex...

[Visual Search]
[Image Info]

Lockheed P2 Neptune Aircraft, , VAH-21, P2 Picture...

[Visual Search]
[Image Info]

Lockheed P2 Neptune Aircraft, , VAH-21, P2 Picture...

[Visual Search]
[Image Info]

Aircraft Base: Military Aircraft, RAF Tornado

[Visual Search]
[Image Info]

Jets, Douglas A4 Skyhawk , Cars, A4

[Visual Search]
[Image Info]

Aircraft Base: Military Aircraft, RAF Tornado

[Visual Search]
[Image Info]

Russian Aviation, Redesignated Su 27ib the Su 34 o...

[Visual Search]
[Image Info]

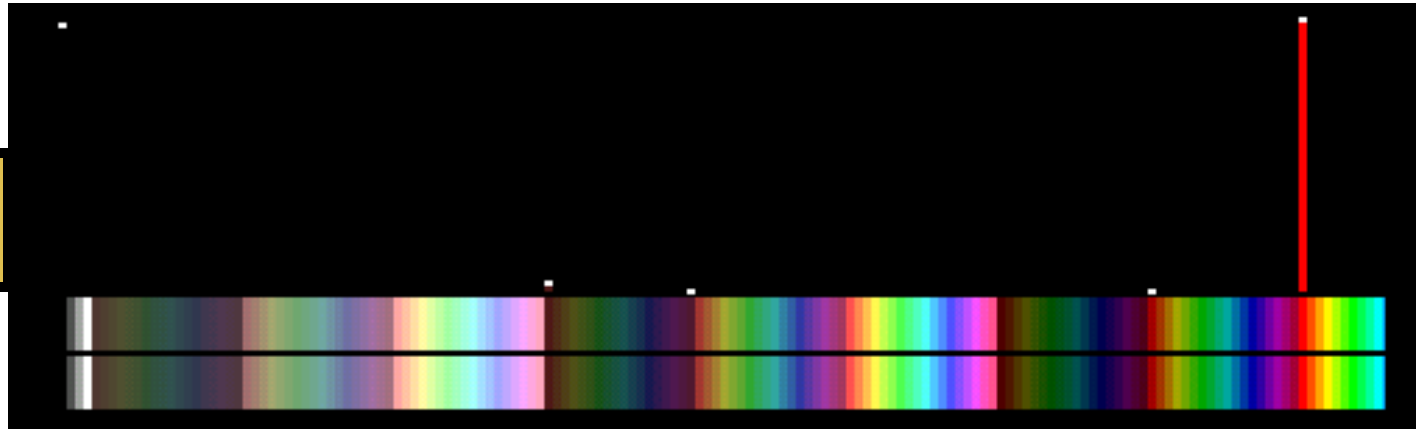
Manitoba Aviation Page, [Photo 3]

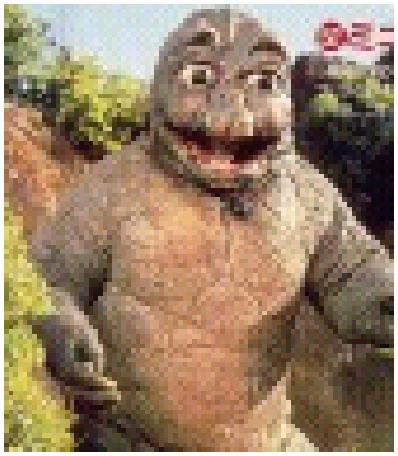
Color Histogram Matching

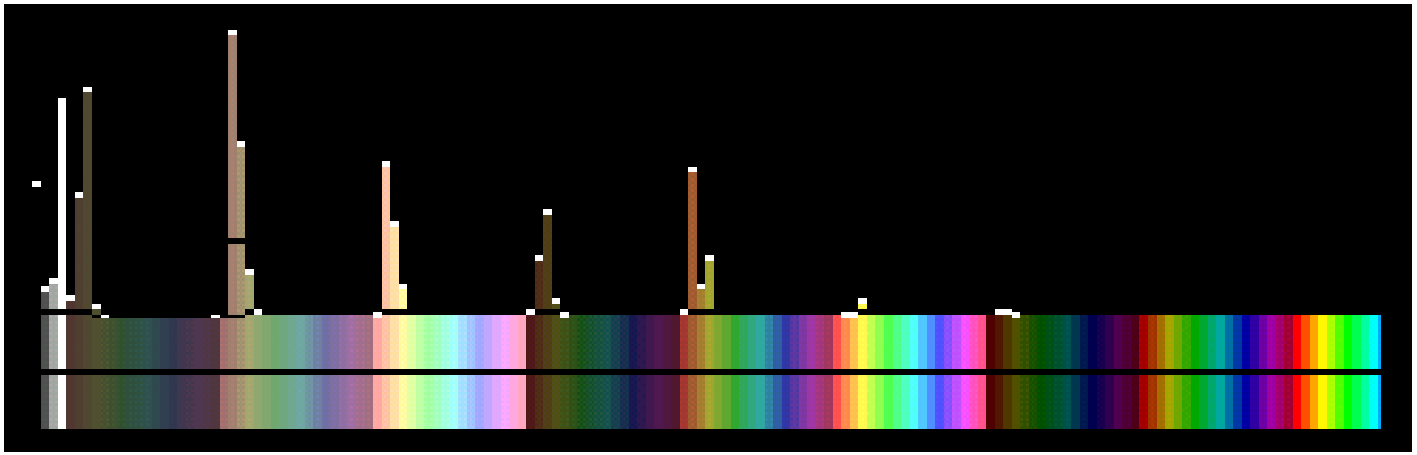
- Represent image as a rectangular pixel raster
  - e.g., 1024 columns and 768 rows
- Represent each pixel as a quantized color
  - e.g., 256 colors ranging from red through violet
- Count the number of pixels in each color bin
  - Produces vector representations
- Compute vector similarity
  - e.g., normalized inner product

http://www.ctr.columbia.edu/webseek/

Color Histogram Example



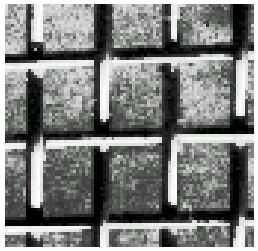


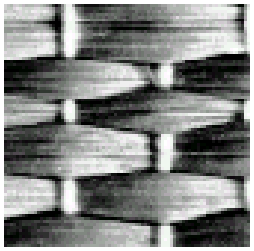


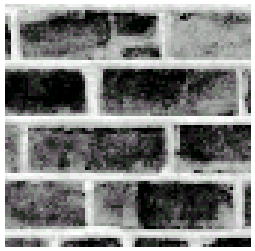
Texture Matching

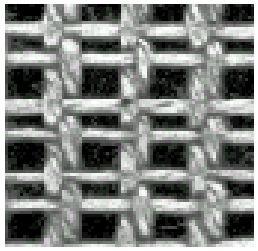
- Texture characterizes small-scale regularity
  - Color describes pixels, texture describes regions
- Described by several types of features
  - e.g., smoothness, periodicity, directionality
- Match region size with image characteristics
  - Computed using filter banks, Gabor wavelets, …
- Perform weighted vector space matching
  - Usually in combination with a color histogram

Texture Test Patterns

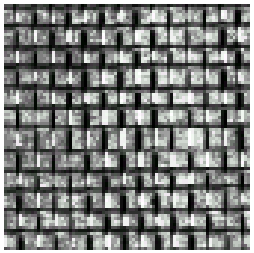

d001

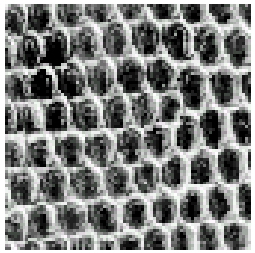

d056

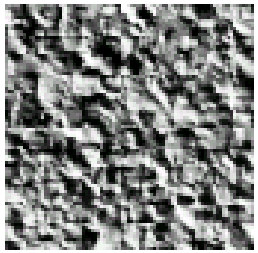

d095

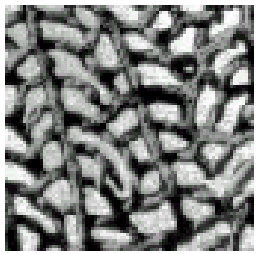

d020


d014

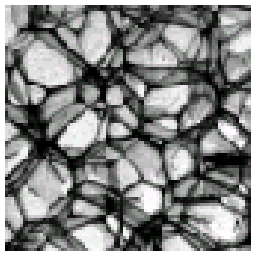

d006

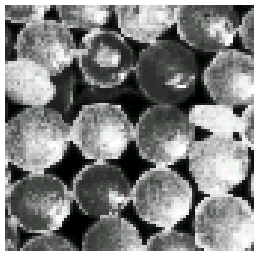

d003

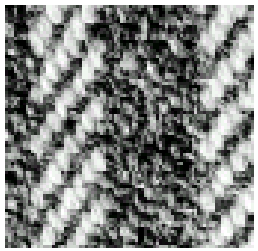

d004

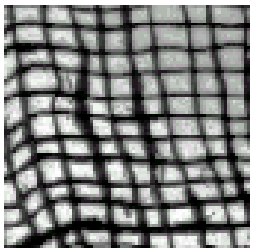

d087

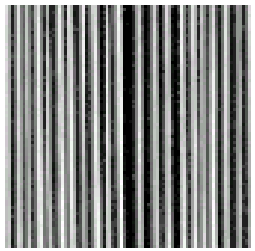

d005


d111


d066


d011


d103


d049

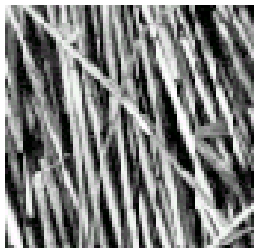
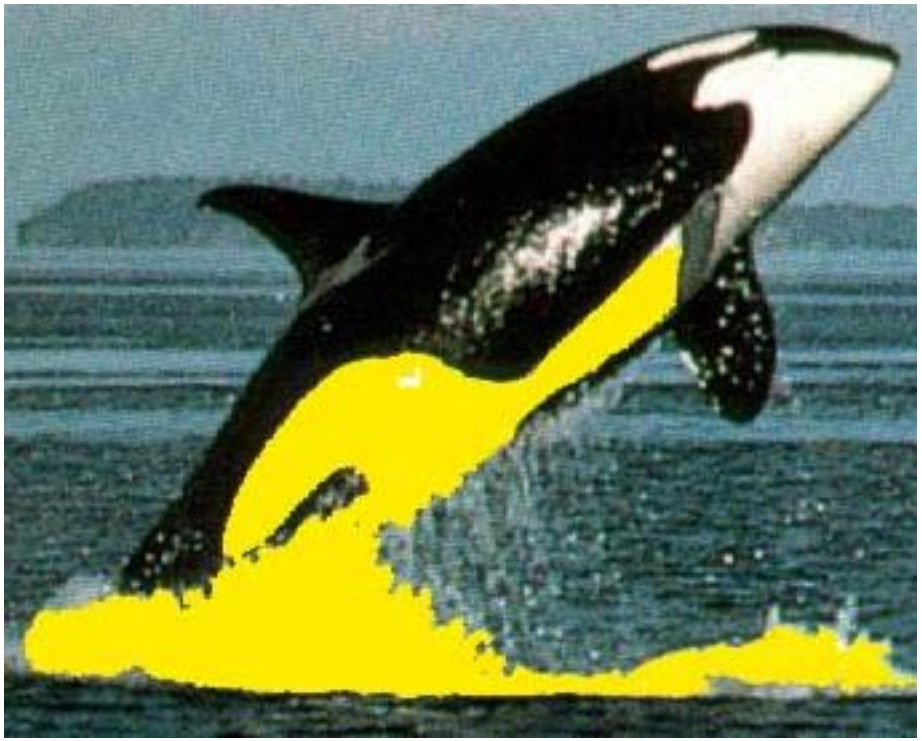

d015

Image Segmentation

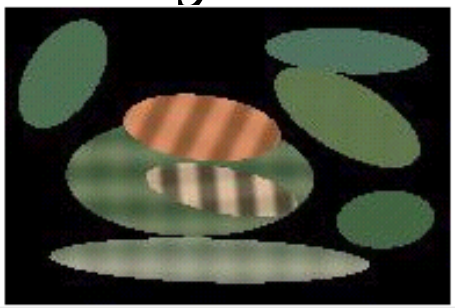
- Global techniques alone yield low precision
  - Color & texture characterize objects, not images
- Segment at color and texture discontinuities
  - Like “flood fill” in Photoshop
- Represent size shape & orientation of objects
  - e.g., Berkeley’s “Blobworld” uses ellipses
- Represent relative positions of objects
  - e.g., angles between lines joining the centers
- Perform rotation- and scale-invariant matching

Flood Fill in Photoshop

- More sophisticated techniques are needed ☺

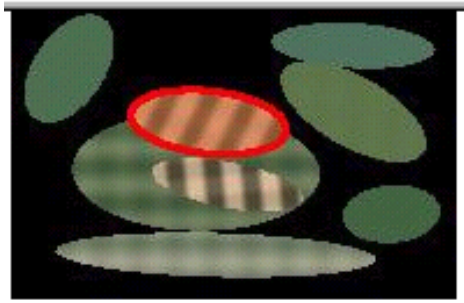


Berkeley Blobworld



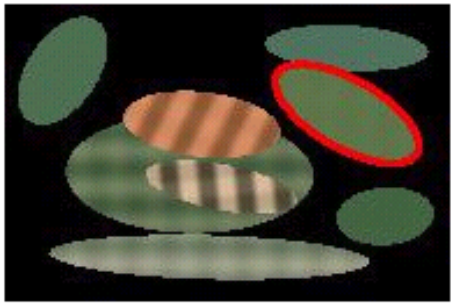
Click on one or two of the blobs in the blobwo
image. Then change the radio buttons to adju
parameter weights. Press one of the Query bu
when you are done.

Full Query | Cheshire Query



	Somewhat Important	Very Important
This blob is:	○	◉

	Not Important	Somewhat Important	Very Important
Color:	○	○	◉
Texture:	○	○	◉
Location:	◉	○	○
Shape/Size:	○	◉	○



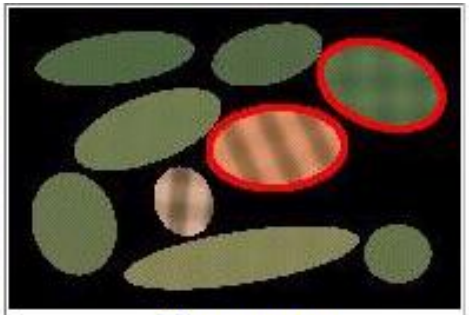
	Somewhat Important	Very Important
This blob is:	◉	○

	Not Important	Somewhat Important	Very Important
Color:	○	○	◉
Texture:	○	◉	○
Location:	◉	○	○
Shape/Size:	◉	○	○

Blob 1 is *to the right of* ○ *to the left of* ○ *above* ○ *below* ○ **Blob 2.**

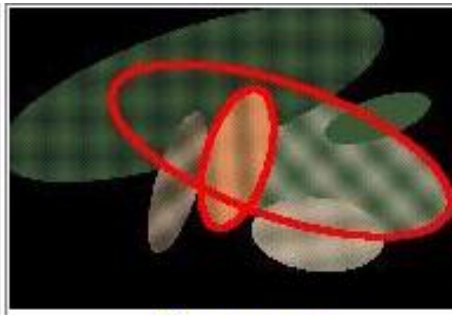
Berkeley Blobworld

108004 (score = 0.86)



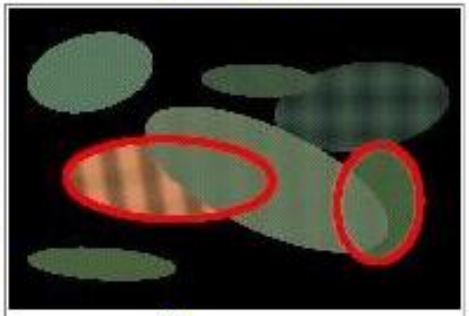
New query

108084 (score = 0.85)



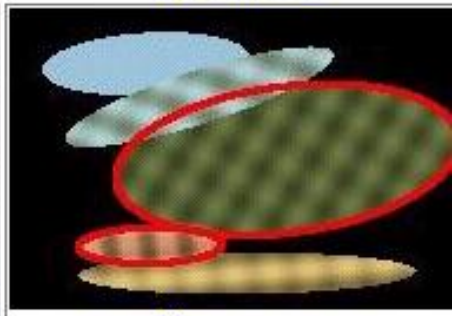
New query

108044 (score = 0.84)



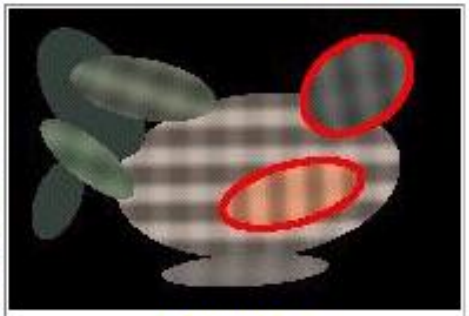
New query

97098 (score = 0.79)



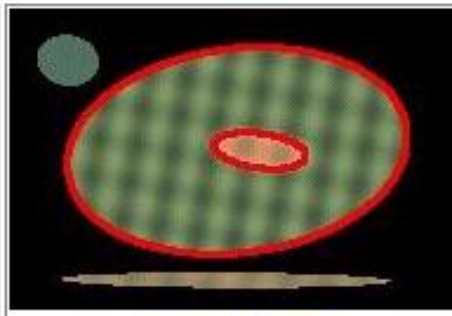
New query

108040 (score = 0.75)



New query

108058 (score = 0.75)



New query

Automated Annotation

door

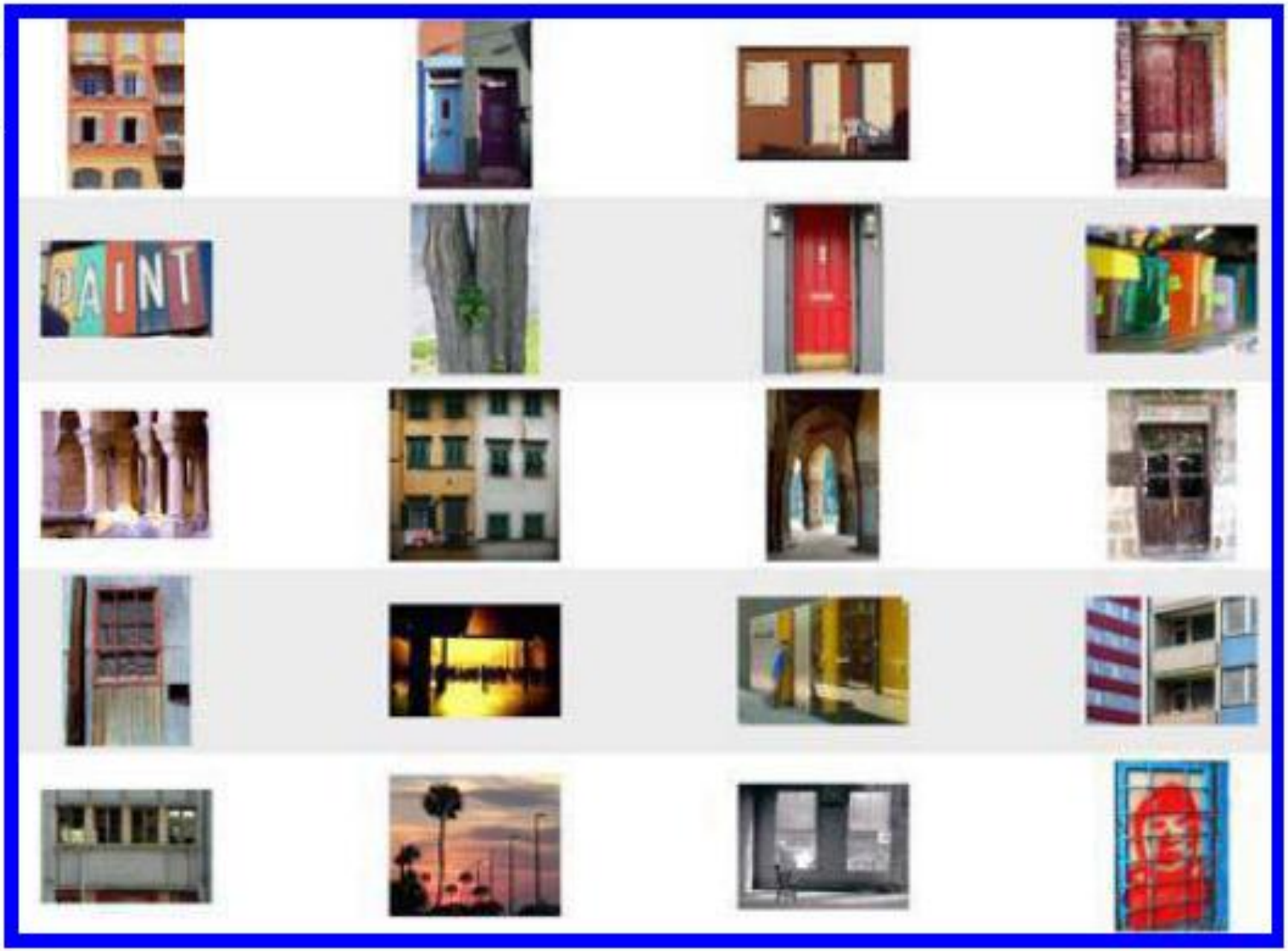


Image Retrieval Summary

- Query
  - Keywords, example, sketch
- Matching
  - Caption text
  - Segmentation
  - Similarity (color, texture, shape)
  - Spatial arrangement (orientation, position)
  - Specialized techniques (e.g., face recognition)
- Selection
  - Thumbnails

Try Some Systems

- Google Image Search (text)
  - http://images.google.com
- IBM QBIC (color, location)
  - http://wwwqbic.almaden.ibm.com/, select Hermitage

Agenda

- Beyond Text, but still language
  - Scanned documents
  - Speech
- Beyond Text, but still information
  - Images
  - ➢Video
- Beyond text to data

Video Structures

- Image structure
  - Absolute positioning, relative positioning
- Object motion
  - Translation, rotation
- Camera motion
  - Pan, zoom, perspective change
- Shot transitions
  - Cut, fade, dissolve, …

Object Motion Detection

- Hypothesize objects as in image retrieval
  - Segment based on color and texture
- Examine frame-to-frame pixel changes
- Classify motion
  - Translation
    - Linear transforms model unaccelerated motion
  - Rotation
    - Creation & destruction, elongation & compression
  - Merge or split

Camera Motion Detection

- Do global frame-to-frame pixel analysis

- Classify the resulting patterns
  - Central tendency -> zoom out
  - Balanced exterior destruction -> zoom in
  - Selective exterior destruction -> pan
  - Coupled rotation and translation -> perspective
    - Coupled within objects, not necessarily across them

Shot-to-Shot Structure Detection

- Create a color histogram for each image
- Segment at discontinuities (cuts)
  - Cuts are easy, other transitions are also detectable
- Cluster representative histograms for each shot
  - Identifies cuts back to a prior shot
- Build a time-labeled transition graph

Shot Classification

- Shot-to-shot structure correlates with genre
  - Reflects accepted editorial conventions
- Some substructures are informative
  - Frequent cuts to and from announcers
  - Periodic cuts between talk show participants
  - Wide-narrow cuts in sports programming
- Simple image features can reinforce this
  - Head-and-shoulders, object size, …

Exploiting Multiple Modalities

- Video rarely appears in isolation
  - Sound track, closed captions, on-screen captions
- This provides synergy, not just redundancy
  - Some information appears in only one modality
- Image analysis complements video analysis
  - Face detection, video OCR

Story Segmentation

- Video often lacks easily detected boundaries
  - Between programs, news stories, etc.
- Accurate segmentation improves utility
  - Too large hurts effectiveness, to small is unnatural
- Multiple segmentation cues are available
  - Genre shift in shot-to-shot structure
  - Vocabulary shift in closed captions
  - Intrusive on-screen text
  - Musical segues

Closed Captions

- Designed for hearing-impaired viewers
  - Speech content, speaker id, non-speech audio
- Weakly synchronized with the video
  - Simultaneously on screen for advance production
  - Significant lag for live productions
- Missing text and significant errors are common
  - Automatic spelling correction can produce nonsense

Aligning Closed Captions

- Speech and closed caption are redundant, but:
  - Each contains different types of errors
  - Each provides unique information
- Merging the two can improve retrieval
  - Start with a rough time alignment
  - Synchronize at points of commonality
    - Speech recognition provides exact timing
  - Use the words from both as a basis for retrieval
    - Learn which to weight more from training data

On-Screen Captions

- On-screen captions can be very useful
  - Speaker names, event names, program titles, …
- They can be very challenging to extract
  - Low resolution, variable background
- But some factors work in your favor
  - Absolutely stable over multiple frames
  - Standard locations and orientations

Video OCR

- Text area detection
  - Look for long thin horizontal regions
    - Bias towards classic text locations by genre
  - Integrate detected regions across multiple frames
- Enhance the extracted text
  - Contrast improvement, interpolation, thinning
- Optical character recognition
  - Matched to the font, if known

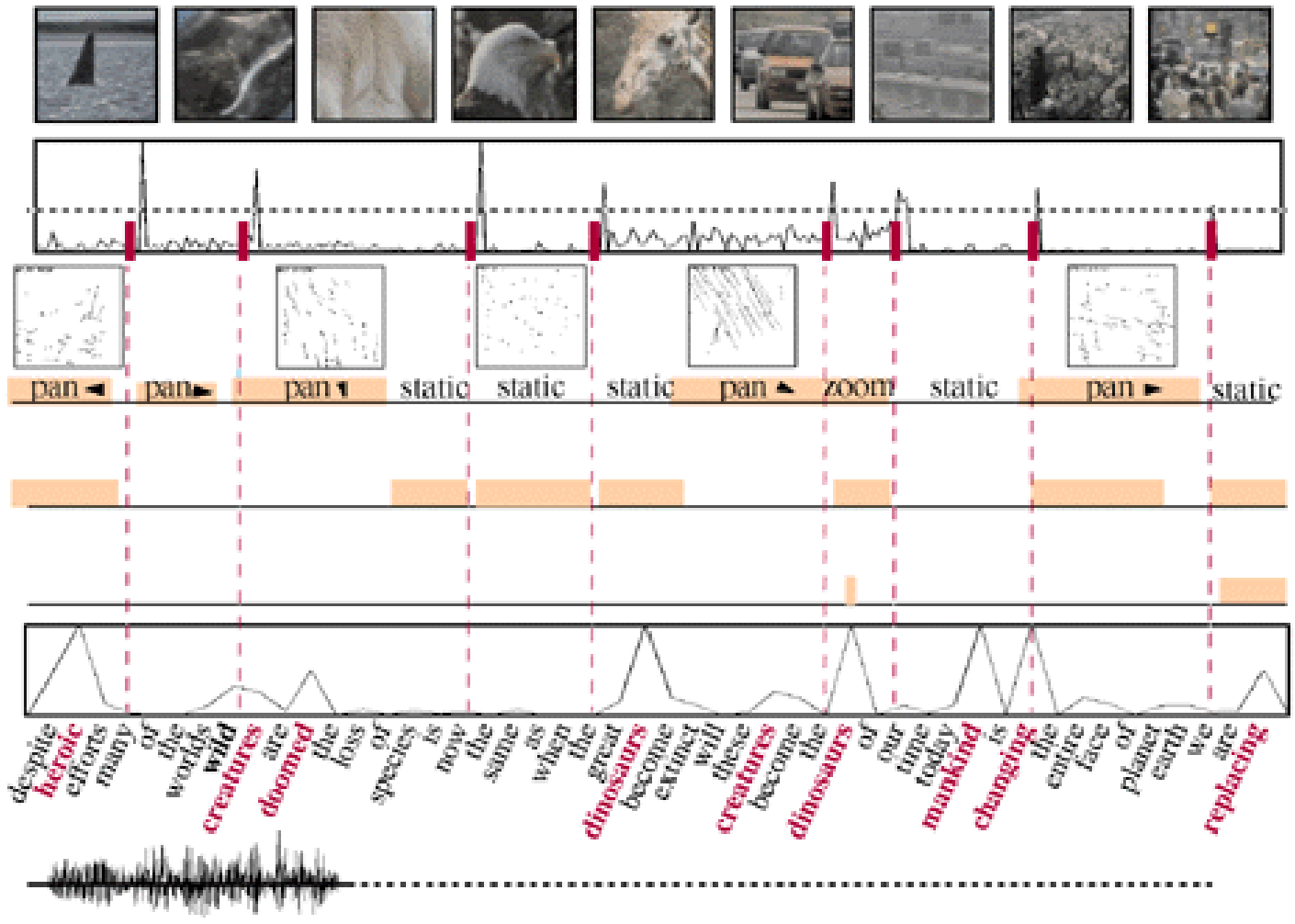
Face Recognition

- Segment from images based on shape
  - Head, shoulders, and hair provide strong cues
- Track across several images
  - Using optical flow techniques
- Select the most directly frontal view
  - Based on eye and cheek positions, for example
- Construct feature vectors
  - “Eigenface” produces 16-element vectors
- Perform similarity matching

Identity-Based Retrieval

- Face recognition and speaker identification
  - Both exploit information that is usually present
  - But both require training data
- On-screen captions provide useful cues
  - Confounded by OCR errors and varied spelling
- Closed captions and speech retrieval help too
  - If genre-specific heuristics are used
    - e.g., announcers usually introduce speakers before cuts

Combined Technologies Integration



Key Frame Extraction

- First frame of a shot is easy to select
  - But it may not be the best choice
- Genre-specific cues may be helpful
  - Minimum optical flow for director's emphasis
  - Face detection for interviews
  - Presence of on-screen captions
- This may produce too many frames
  - Color histogram clusters can reveal duplicates

CMU Informedia DVLS v. 1.06

File Edit Navigate Options Video Window Help Comments!

tell me about the evolution of species

Clear All! More Options... OR: Search

The results set shows the best 12 of 628 matches on any of "tell me about the evolution of species."

Click on a word to focus on it. Press the shift key while clicking to have multi-word focus.

Search Results

Darwin observed Galapagos and developed theor...

Resume > |<< Prev Hit 00047340 Next Hit >>| 0:00:42

<< Prev Para. << All >> Next Para. >> 0:57:11

When Charles Darwin came here 150 years ago he was puzzled by the differences between similar animals from

Salient Stills Abstracts

- Composite images that capture several scenes
  - And convey a sense of space, time, and/or motion
- Exploits familiar metaphors
  - Time exposures, multiple exposures, strobe, …
- Two stages
  - Modeling (e.g., video structure analysis)
  - Rendering
    - Global operators do time exposure and variable resolution
    - Segmentation supports production of composite frames

Storyboards

- Spatial arrangement of still images
  - Linear arrangements depict temporal evolution
    - Overlapped depictions allow denser presentations
  - Graph can be used to depict video structure
    - But temporal relationships are hard to capture
- Naturally balances overview with detail
  - Easily browsed at any level of detail
- Tradeoff between detail and complexity
  - Further limited by image size and resolution

Static Filmstrip Abstraction

Slide Shows

- Flip through still images in one spot
  - At a rate selected by the user
- Conserves screen space
  - But it is hard to process several simultaneously
- Several variations possible
  - Content-sensitive dwell times
  - Alternative frame transitions (cut, dissolve, …)

Full Motion Extracts

- Extracted shots, joined by cuts
  - The technique used in movie advertisements
- Conveys more information using motion
  - Optionally aligned with extracted sound as well
- Hard to build a coherent extract
  - Movie ads are constructed by hand

Agenda

- Beyond Text, but still language
  - Scanned documents
  - Speech
- Beyond Text, but still information
  - Images
  - Video

➢ Beyond text to data

Uses of Relational Databases

- Relatively static “relational” data
  - Organization charts
  - Phone directories
- Transaction processing
  - Financial
  - B2B
  - B2C

Database Implementation

- DBMS
- Normalized tables
  - Organized to minimize implementation
- Denormalized tables
  - Organized to support access
- SQL
- File system
- CSV
- Programming language

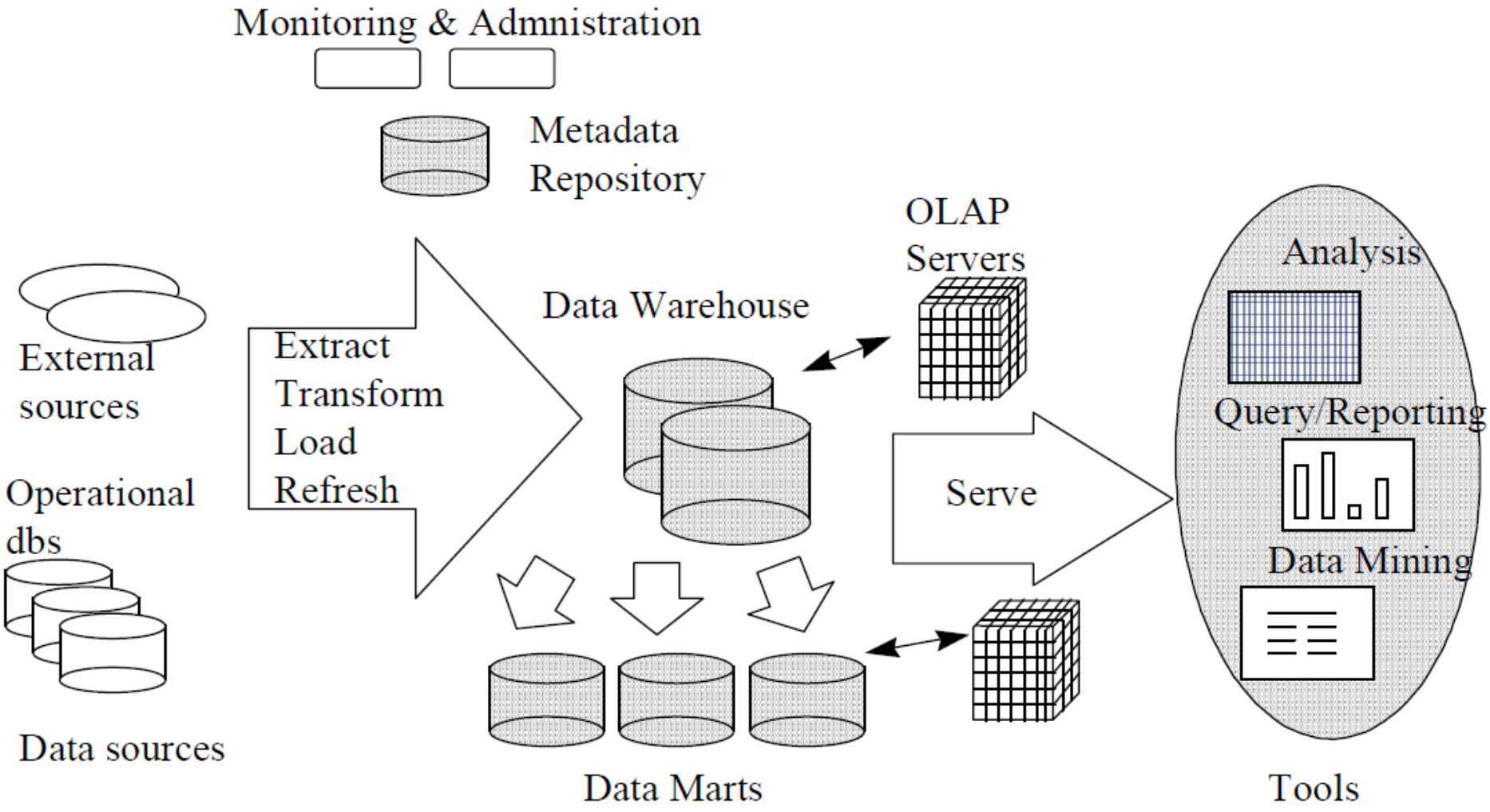
Operational Systems

- Data exchange
- Forms
- Views
- Reports

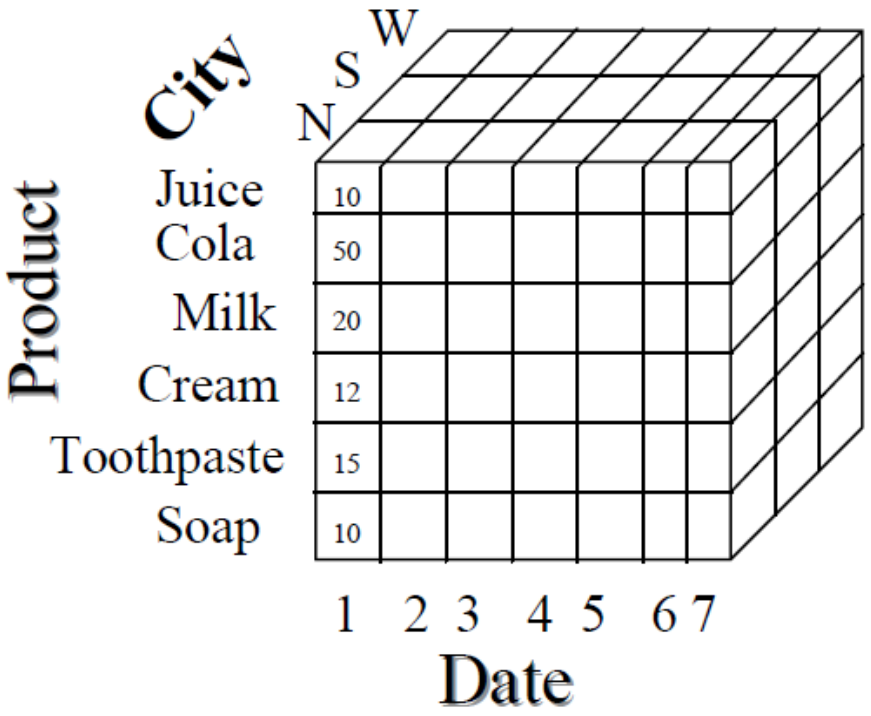
Data Warehouse

- Non-operational “strategic” data resource
  - Organized for analysis, not operations
- Data harvested from operational systems
- OLAP
- Data mining
  - Association rules

Data Warehousing

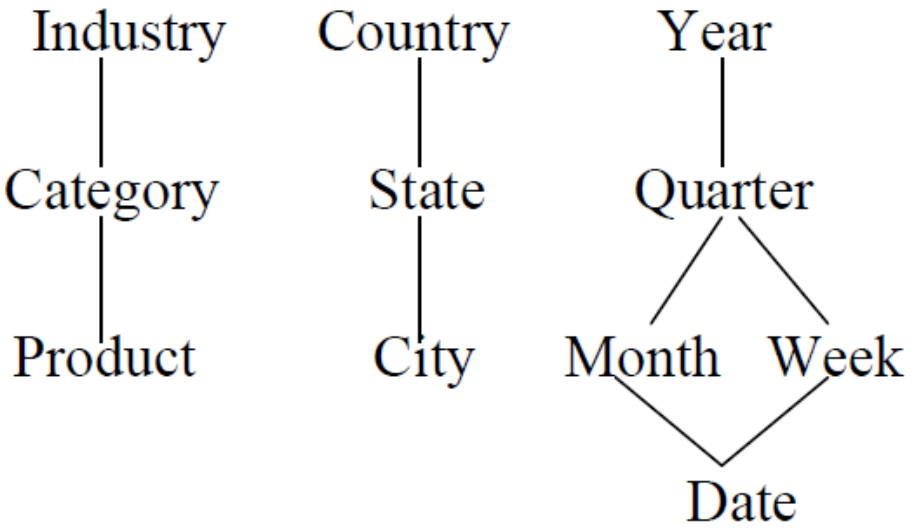


Online Analytical Processing (OLAP)



Dimensions: Product, City, Date

Hierarchical summarization paths



Sedona Database Principles

1. Absent a specific showing of need or relevance, a requesting party is entitled only to database **fields** that contain relevant information, not the entire database in which the information resides or the underlying database application or database engine.

2. Due to differences in the way that information is stored or programmed into a database, not all information in a database may be equally accessible, and a party's request for such information must be analyzed for relevance and proportionality.

3. Requesting and responding parties should use empirical information, such as that generated from test queries and pilot projects, to ascertain the burden to produce information stored in databases and to reach consensus on the scope of discovery.

4. A responding party must use reasonable measures to validate ESI collected from database systems to ensure completeness and accuracy of the data acquisition.

5. Verifying information that has been correctly exported from a larger database or repository is a separate analysis from establishing the accuracy, authenticity, or admissibility of the substantive information contained within the data.

6. The way in which a requesting party intends to use database information is an important factor in determining an appropriate format of production.