KNOWLEDGE MEDIA
KMi
INSTITUTE

Multimedia Information Retrieval

1 What is multimedia information retrieval?
2 Basic Multimedia Search Technologies
3 Evaluation of MIR Systems
4 Added Value – user interaction, visualisation and the MIR research landscape

KNOWLEDGE MEDIA INSTITUTE

Multimedia Information Retrieval

KNOWLEDGE MEDIA INSTITUTE

Evaluation

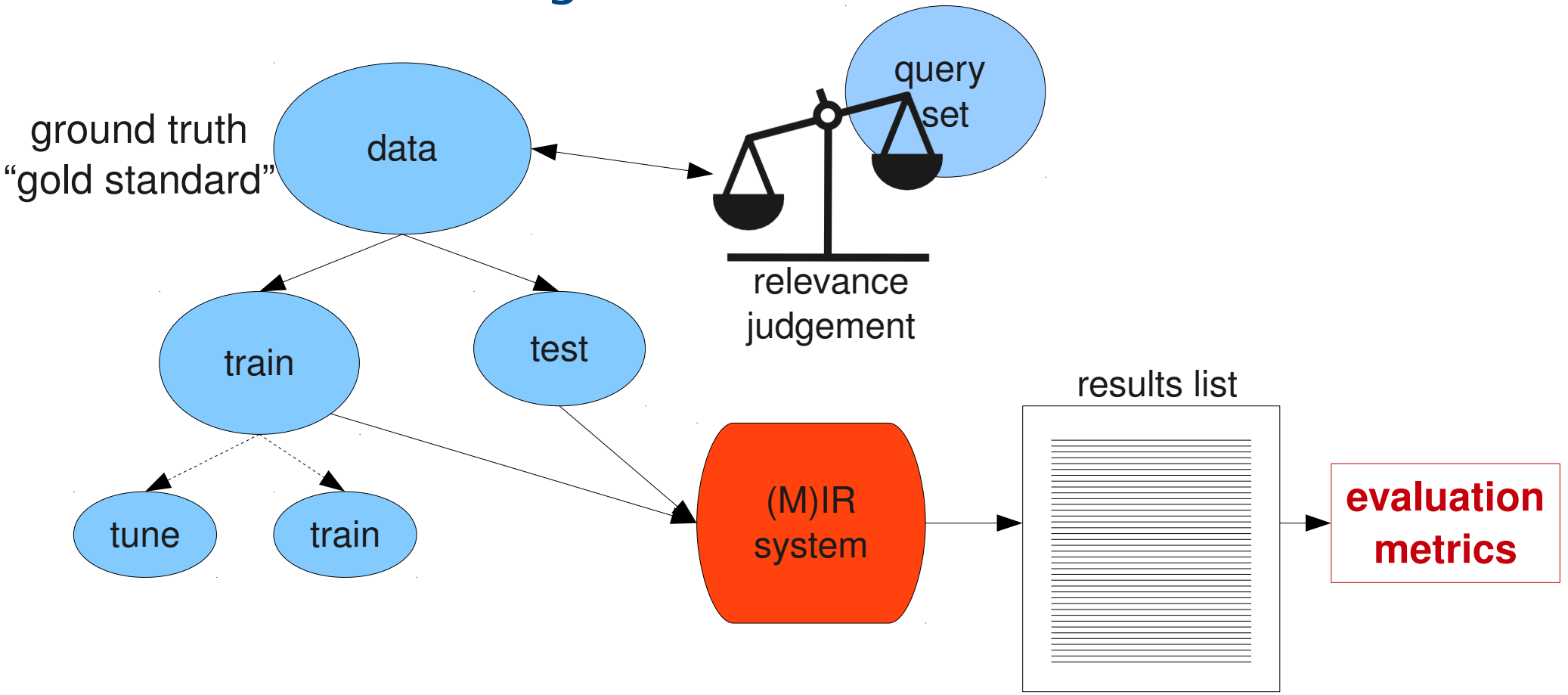
How do we know if our MIR system is effective?

Why do we care about quantifying the performance?

“If you can not measure it, you can not improve it.” – Lord Kelvin

Information Retrieval

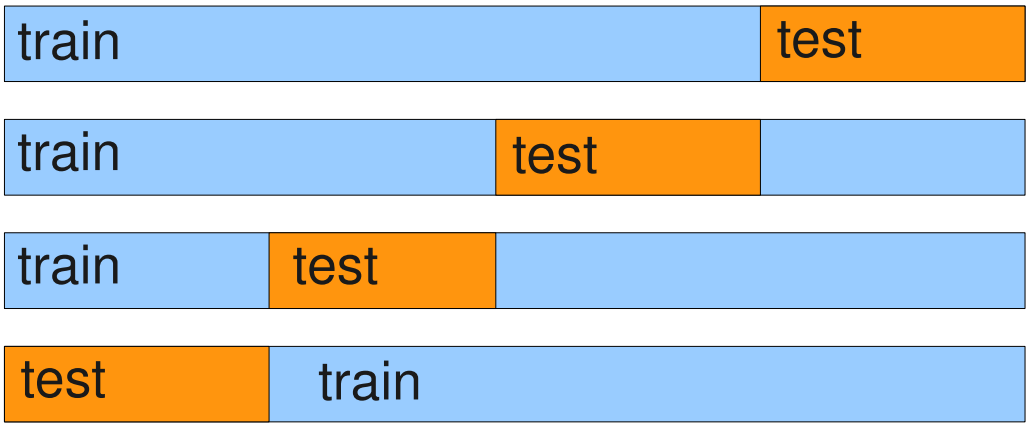
"Cranfield Paradigm"



William Webber, 'When did the Cranfield tests become the "Cranfield paradigm"?'
http://blog.codalism.com/?p=817

Small, unbalanced data?

Cross-validation



Randomise data and divide
Train-test 4 times
Average all metrics
4-fold cross-validation

Extreme is Leave-One-Out : test size = 1

KNOWLEDGE MEDIA KMi INSTITUTE

Relevance?

Find me pictures of triumph

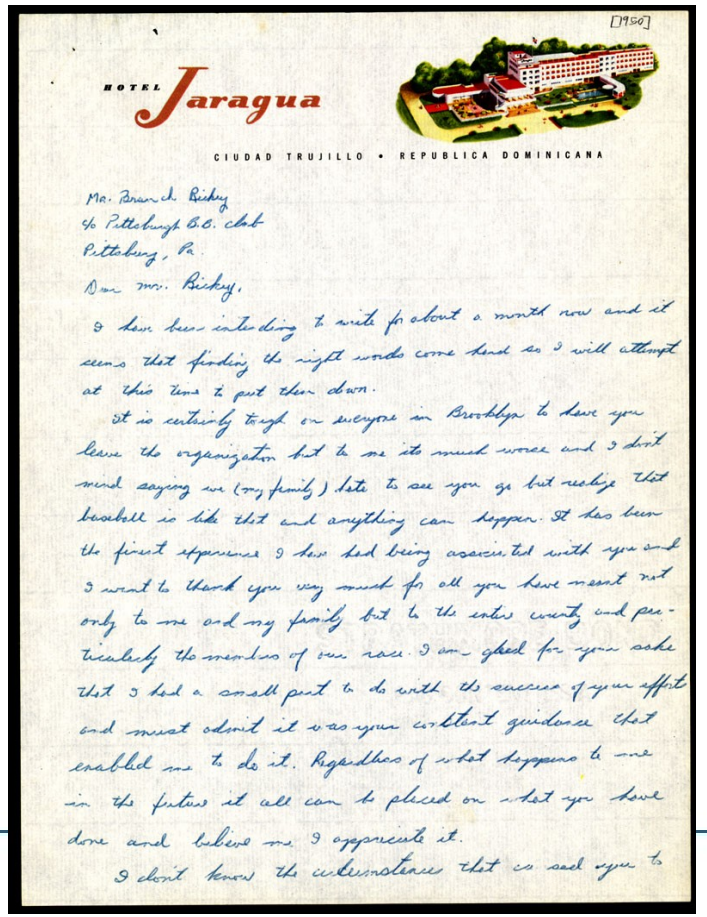
*http://www.flickr.com/photos/ricardodiaz/

The Open University

KNOWLEDGE MEDIA KMi INSTITUTE

Exercise

Find shots of printed, typed, or handwritten text, filling more than half of the frame area



HOTEL Jaragua

CIUDAD TRUJILLO • REPUBLICA DOMINICANA

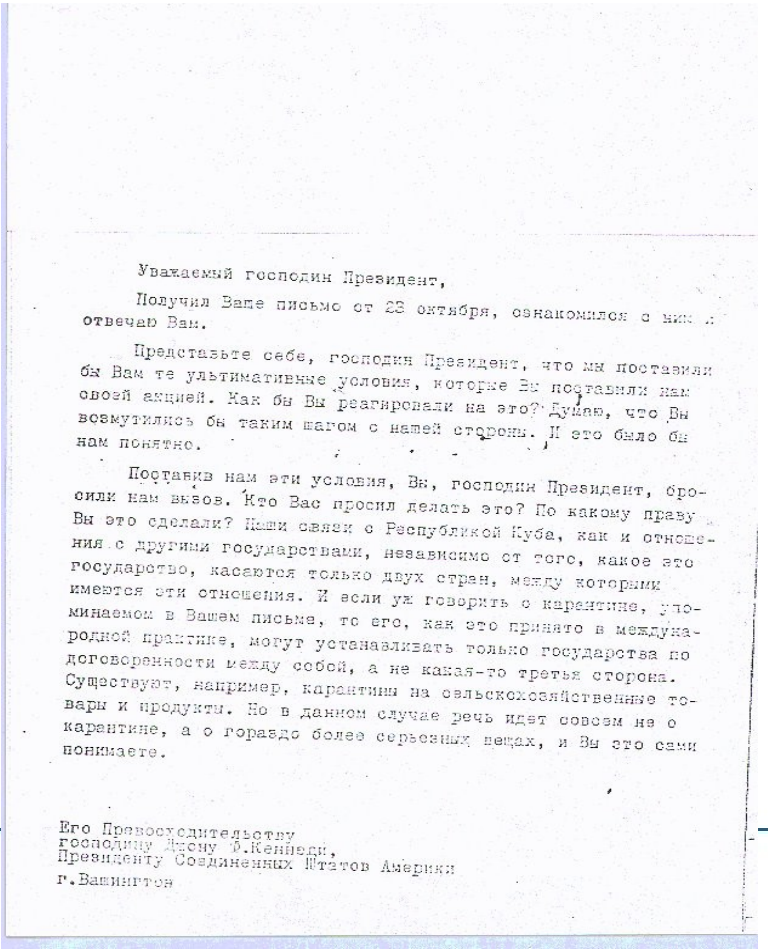
Mr. Brandt Bixby
c/o Pittsburgh B.C. Club
Pittsburg, Pa.

Dear Mr. Bixby,

I have been intending to write for about a month now and it seems that finding the right words come hard so I will attempt at this time to put them down.

It is entirely tough on everyone in Brooklyn to have you leave the organization but to me its much worse and I don't mind saying we (my family) hate to see you go but realize that baseball is like that and anything can happen. It has been the finest experience I have had being associated with you and I want to thank you very much for all you have meant not only to me and my family but to the entire county and particularly the members of our race. I am glad for your sake that I had a small part to do with the success of your effort and must admit it was your excellent guidance that enabled me to do it. Regardless of what happens to me in the future it all can be placed on what you have done and believe me I appreciate it.

I don't know the circumstances that is said you to



Уважаемый господин Президент,

Получил Ваше письмо от 23 октября, ознакомился с ним и отвечаю Вам.

Представьте себе, господин Президент, что мы поставили бы Вам те ультимативные условия, которые Вы поставили нам своей акцией. Как бы Вы реагировали на это? Думаю, что Вы возмутились бы таким шагом с нашей стороны. И это было бы нам понятно.

Поставив нам эти условия, Вы, господин Президент, бросили нам вызов. Кто Вас просил делать это? По какому праву Вы это сделали? Наши связи с Республикой Куба, как и отношения с другими государствами, независимо от того, какое это государство, касаются только двух стран, между которыми имеются эти отношения. И если уж говорить о карантине, упоминаемом в Вашем письме, то его, как это принято в международной практике, могут устанавливать только государства по договоренности между собой, а не какая-то третья сторона. Существуют, например, карантины на сельскохозяйственные товары и продукты. Но в данном случае речь идет совсем не о карантине, а о гораздо более серьезных вещах, и Вы это сами понимаете.

Его Превосходительству
господину Джону Ф.Кеннеди,
Президенту Соединенных Штатов Америки
г.Вашингтон

Metrics

	Relevant	**Irrelevant**
Retrieved	True Positive (tp)	False Positive (fp)
Not Retrieved	False Negative (fn)	True Negative (tn)

Precision (P) = fraction retrieved that are relevant

$P = tp/(tp+fp)$

Recall (R) = fraction relevant that are retrieved

$R = tp/(tp+fn)$

Precision or Recall?

What about accuracy?

Accuracy = (tp+tn)/(tp+fp+fn+tn)

Is precision or recall more useful/important

if I'm doing a web search on Gold Coast accommodation?

if I'm a paralegal researching case precedence?

How could I make a system with 100% recall?

F_1-measure (weighted harmonic mean of P & R)

$$F_1 = \frac{2 \cdot precision \cdot recall}{(precision + recall)}$$

KNOWLEDGE MEDIA INSTITUTE

Exercise

An IR system returns 8 relevant documents and 10 irrelevant documents. There are a total of 20 relevant documents in the collection. Calculate the precision and recall.

Exercise

An IR system returns 8 relevant documents and 10 irrelevant documents. There are a total of 20 relevant documents in the collection. Calculate the precision and recall.

tp = 8	fp = 10
fn = 12	tn = (unknown)

P = tp/(tp+fp) = 8/(8+10) = 8/18 = 0.44

R = tp/(tp+fn) = 8/(8+12) = 8/20 = 0.40

F_1-measure would be 2x0.44x0.40/(0.44+0.40) = 0.42

KNOWLEDGE MEDIA KMi INSTITUTE

Ranked Retrieval

Which is better? There are 5 relevant documents to be found.

System A

1. Relevant
2. Relevant
3. Irrelevant
4. Irrelevant
5. Relevant
6. Relevant

Precision = 4/6 = 0.66
Recall = 4/5 = 0.80

System B

1. Relevant
2. Irrelevant
3. Relevant
4. Relevant
5. Relevant
6. Irrelevant

Precision = 4/6 = 0.66
Recall = 4/5 = 0.80

KNOWLEDGE MEDIA INSTITUTE

Ranked Retrieval Metrics

Precision @ N

Precision/Recall graphs

Mean Average Precision

KNOWLEDGE MEDIA KMi INSTITUTE

Ranked Retrieval

Which is better? There are 5 relevant documents to be found.

System A	System B	
1. Relevant	1. Relevant	P@1
2. Relevant	2. Irrelevant	P@2
3. Irrelevant	3. Relevant	P@3
4. Irrelevant	4. Relevant	P@4
5. Relevant	5. Relevant	P@5
6. Relevant	6. Irrelevant	

Precision/Recall Curve

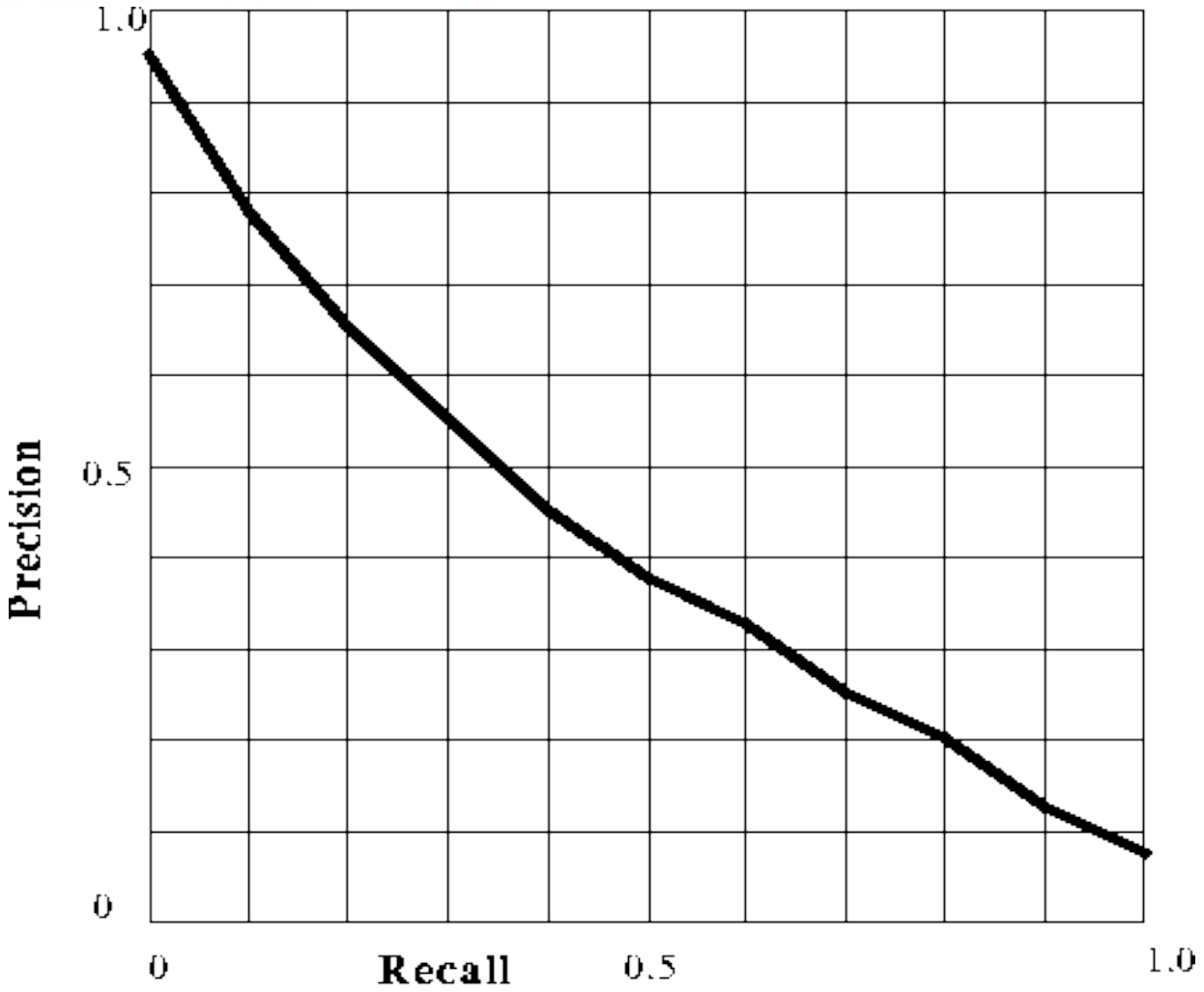


Fig. 1. A typical precision-recall graph

KNOWLEDGE MEDIA KMi INSTITUTE

(Mean) Average Precision

System A

1. Relevant P = 1
2. Relevant P = 1
3. Irrelevant -
4. Irrelevant -
5. Relevant P = 0.6
6. Relevant P = 0.67

(1+1+0.6+0.67)/4 = 0.82

System B

1. Relevant P = 1
2. Irrelevant -
3. Relevant P = 0.67
4. Relevant P = 0.75
5. Relevant P = 0.8
6. Irrelevant -

(1+0.67+0.75+0.8)/4 = 0.69

KNOWLEDGE MEDIA INSTITUTE

Ranked Retrieval

Which is better? There are 5 relevant documents to be found.

System A

1. Relevant
2. Relevant
3. Irrelevant
4. Irrelevant
5. Relevant
6. Relevant

AP = 0.82

System B

1. Relevant
2. Irrelevant
3. Relevant
4. Relevant
5. Relevant
6. Irrelevant

AP = 0.69

KNOWLEDGE MEDIA INSTITUTE

Exercise

Use the results (exercises/evaluation/) from 2 image search engines and calculate the performance. Which is better?

Spreadsheet

KNOWLEDGE MEDIA INSTITUTE

The Dark Side of Evaluation ...

- Overfitting to limited training data → unbalanced, fragile system
- Unrealistic training data
- Difficulty in finding training data
- Comparison and competition
- Numbers not users

KNOWLEDGE MEDIA INSTITUTE

Evaluation Campaigns

TRECVID

ImageCLEF

MediaEval

MIREX

KNOWLEDGE MEDIA INSTITUTE

TREC Video retrieval conferences

Organised by NIST with support from other U.S. government agencies - http://www-nlpir.nist.gov/projects/trecvid/

Objective is to encourage research in information retrieval by:

Providing a **large** test collection.

Uniform **scoring** procedures.

Forum for organizations interested in **comparing** their results.

Tasks:

Shot boundary detections (retired)

High-level feature extraction (semantic annotation)

Search (interactive, manually-assisted or fully automated)

Rushes summarisation

KNOWLEDGE MEDIA INSTITUTE

TRECVID's dirty secret

In the first few years of TRECVID video retrieval was best done with “text only”

Image analysis did not help in early years

BUT situation has changed!

Combination of weak classifiers to corroborate evidence

The number of visual concepts has increased; see, eg, LSCOM

TRECVid

TRECVid example queries

"Find shots of a road taken from a moving vehicle through the front window"

"Find shots of a person talking behind a microphone"

"Find shots of a street scene at night"

KNOWLEDGE MEDIA INSTITUTE

ImageCLEF

CLEF = Cross Language Evaluation Forum

Process is modelled from TREC

ImageCLEF started in 2003

Tasks:

- Image retrieval (queries in different languages)
- Medical Image Annotation
- Annotation of photographs
- Geographic retrieval (GeoCLEF)
- Video retrieval (VideoCLEF/MediaEval)

KNOWLEDGE MEDIA INSTITUTE

Search Engine Quality?

System issues

- Indexing speed
- Scalability
- Robustness
- Query expressiveness

User issues

- Diversity, Responsiveness
- “happiness” ?
- The interface vs IR performance

KNOWLEDGE MEDIA INSTITUTE

Multimedia Information Retrieval