Creative Computing II

Christophe Rhodes
c.rhodes@gold.ac.uk

Autumn 2010, Wednesdays:
10:00–12:00: RHB307 & 14:00–16:00: WB316
Winter 2011, Wednesdays:
10:00–12:00: RHB307 & 14:00–16:00: WB316

Multimedia Information Retrieval

Information Retrieval:

- name given to general field of retrieving information in response to a **query**;
- Van Rijsbergen, C.J. *Information Retrieval*.

Nature of query and retrieval:

- query: specificity scale;
- retrieval: only exact, or also approximate matches;
- domain and range of retrieval process.

Multimedia Information Retrieval

Specificity

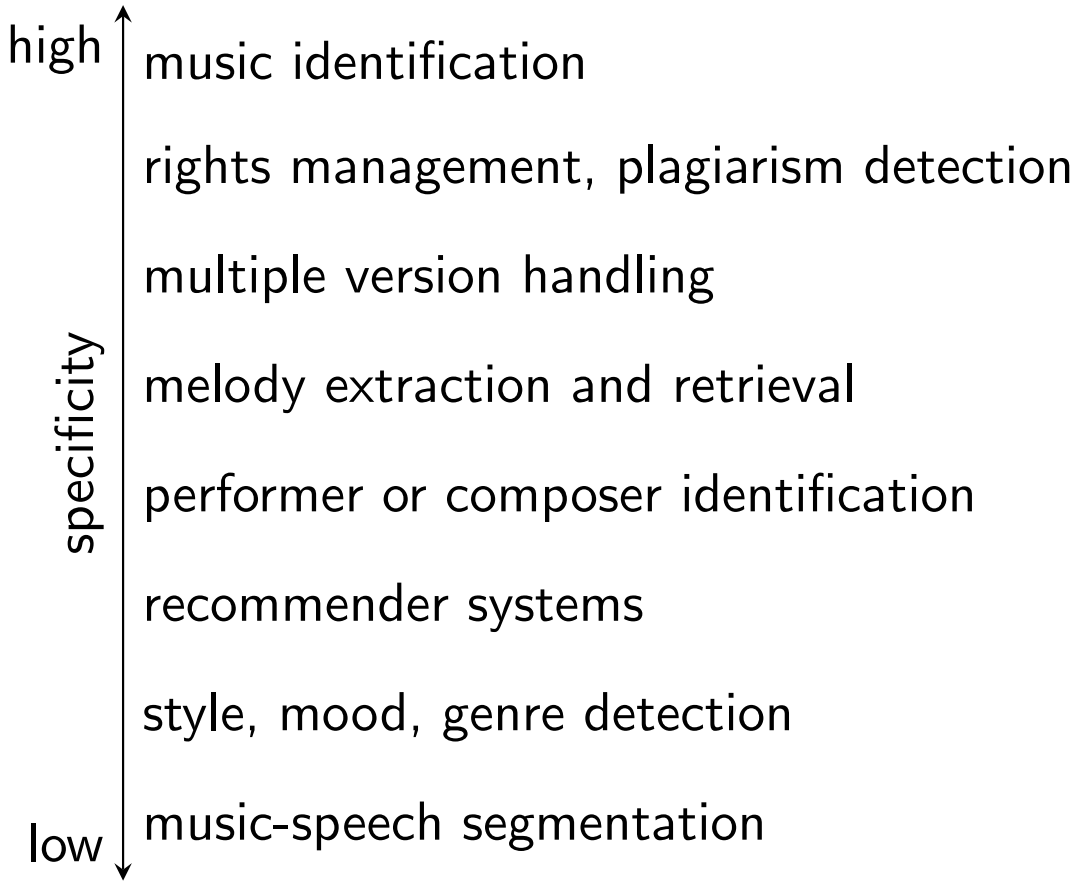
How **specific** is the query? Examples:

- 'the fourth song on the *Sgt Pepper* album';
- '*Hallelujah*';
- 'something bluesy';
- 'a track I would like to listen to now'.

Multimedia Information Retrieval

Specificity

Music Information Retrieval Specificity



After *A Similarity Scale for Content-Based Music IR*, D. Byrd

Multimedia Information Retrieval

Exactness

How stringent are our requirements on the retrieval?

- 'the fourth song on the *Sgt Pepper* album';
- '*Hallelujah*';
- 'something bluesy';
- 'a track I would like to listen to now'.

Are only **exact** matches acceptable, or are **approximate** matches good enough too?

Multimedia Information Retrieval

Domain and Range

What are we trying to retrieve information from?

- database (**corpus**) of materials?
- album?
- track?
- 100ms of audio?

What information are we trying to retrieve?

- album?
- track?
- musical information (*e.g.* key)?
- metadata?

Multimedia Information Retrieval

Fingerprinting

Problem statement: produce a unique identifier for a piece of multimedia that is (reasonably) invariant to distortion.

$$\text{media item} \rightarrow \text{perceptual hash}$$

$$\text{item} \approx \text{item}' \implies \text{hash} = \text{hash}'$$

Purposes:

- duplicate identification;
- rights management;
- directed sales.

e.g. Shazam, pHash.

Multimedia Information Retrieval

Text-based search

Problem statement: given some textual metadata, retrieve media items to which that metadata applies.

$$\text{textual metadata} \rightarrow \text{media item}^*$$

Purposes:

- content delivery;
- personal collection organization;
- media navigation and discovery.

e.g. music player search, Amazon

Multimedia Information Retrieval

Similarity

Problem statement: find items from a database which are 'similar' in some way to a query.

$$\text{media item} \rightarrow \text{media item}^*$$

$$\text{item} \approx \text{item}'$$

Purposes:

- media discovery;
- rights management.

e.g. SoundBite, audioDB

Multimedia Information Retrieval

Music Similarity

What does ‘similar’ mean?

- cover song?
- remix? mashup?
- same key? same genre? style?
- same structure? same artist?

Different things to different people (at different times).

Multimedia Information Retrieval

Textual Features

- textual metadata;
- collaborative filtering.

Multimedia Information Retrieval

Textual Features

- textual metadata;
- collaborative filtering.

Textual feature treatment techniques:

- **stopword** or **noise word** removal;
- **stemming**;
- distance measures.

Multimedia Information Retrieval

Textual Features

- textual metadata;
- collaborative filtering.

Textual feature treatment techniques:

- **stopword** or **noise word** removal;
- **stemming**;
- distance measures.

Common search strategy:

- term-frequency–inverse-document-frequency

Multimedia Information Retrieval

Textual Features

stopword removal:

- some common words are not useful in an index.
- *e.g.* 'the', 'a', 'but', 'who', 'I'
- these words are typically removed prior to construction of an index (or ignored in distance measures if there is no index).

Stemming:

- many words come in different forms
- verbs: conjugation;
- nouns: pluralization;
- adverb / adjective duality.

Identify the **stem** of the word, so that all variants are findable. (cf. **Porter stemmer**)

Multimedia Information Retrieval

Textual Distance Measures

Identity measure:

- if the two words are exactly the same, their distance is 0;
- otherwise, the distance between them is 1.

Multimedia Information Retrieval

Textual Distance Measures

Identity measure:

- if the two words are exactly the same, their distance is 0;
- otherwise, the distance between them is 1.

$d \leftarrow d_{Identity}(x, y)$

$l_x \leftarrow \text{length}(x)$; $l_y \leftarrow \text{length}(y)$
if $\neg(l_x = l_y)$ **then**
  $d \leftarrow 1$
else
  $d \leftarrow 0$
  for i from 1 to l_x **do**
    if $\neg(x_i = y_i)$ **then**
      $d \leftarrow 1$
      return
    end if
  end for
end if

This distance computation is $O(L)$ in time for strings of length L.

Multimedia Information Retrieval

Textual Distance Measures

The Identity distance measure is too specific for most uses:

- usually want to have some tolerance (*e.g.* for misspellings)
- all non-identity pairs at the *same* distance.

Examples:

$$d(\texttt{choose}, \texttt{choose}) = 0$$

$$d(\texttt{choose}, \texttt{chose}) = 1$$

$$d(\texttt{choose}, \texttt{chives}) = 1$$

$$d(\texttt{professor}, \texttt{proffessor}) = 1$$

$$d(\texttt{professor}, \texttt{cabbage}) = 1$$

Multimedia Information Retrieval

Textual Distance Measures

Hamming distance:

- if the two words have the same length, then their Hamming distance is the number of positions at which they differ;
- if the two words have different lengths, the Hamming distance is undefined.

$d \leftarrow d_{Hamming}(x, y)$

$l_x \leftarrow \text{length}(x)$; $l_y \leftarrow \text{length}(y)$
if $\neg(l_x = l_y)$ **then**
  $d \leftarrow \perp$
else
  $d \leftarrow 0$
  for i from 1 to l_x **do**
    if $\neg(x_i = y_i)$ **then**
      $d \leftarrow d + 1$
    end if
  end for
end if

This distance computation is $O(L)$ in time for strings of length L.

Multimedia Information Retrieval

Textual Distance Measures

The Hamming distance is not ideal for natural language string:

- many misspellings change a word's length; doesn't model common ways of making mistakes.
- (useful in other contexts: particularly bit strings)

$$d(\texttt{choose}, \texttt{choose}) = 0$$

$$d(\texttt{choose}, \texttt{chose}) = \perp$$

$$d(\texttt{choose}, \texttt{chives}) = 4$$

$$d(\texttt{professor}, \texttt{proffessor}) = \perp$$

$$d(\texttt{professors}, \texttt{proffessor}) = 5$$