http://www.xkcd.com/655/
Audio Retrieval

David Kauchak
cs160
Fall 2009

Thanks to Doug Turnbull for some of the slides
Administrative

- CS Colloquium vs. Wed. before Thanksgiving
  😊
producers

8M artists

150M songs

1M downloads/month

consumers

250M iPods

9B songs

93M americans
Audio Index construction

Audio files to be indexed

- wav
- midi
- mp3

audio preprocessing

indexer

Index

may be keyed off of text

may be keyed off of audio features
Audio retrieval

Systems differ by what the query input is and how they figure out the result.
Song identification

Given an audio signal, tell me what the “song” is

- Examples:
  - Query by Humming
  - 1-866-411-SONG
  - Shazam
  - Bird song identification
  - ...

Song identification

- How might you do this?

Query by humming

“song” name
Song identification

“song” name
Song similarity

Find the songs that are most similar to the input song

- Examples:
  - Genius
  - Pandora
  - Last.fm
Song similarity

- How might you do this?

IR approach

rank by cosine sim
Song similarity: collaborative filtering

What might you conclude from this information?
Songs using descriptive text search

- Examples:
  - Very few commercial systems like this …
Meerkat

Type a tag to begin

ex. pop, cheerful, urban
Meerkat

Type a tag to begin

bluegrass

ex. pop, cheerful, urban

Add Tag
Meerkat

Spooky Lane by Sam Bush

Uncle Pen by Bill Monroe & His Bluegrass Boys

I'll Never Shed Another Tear by The Osborne Brothers

'Spooky Lane' Tags:

- country influences
- theme time radio hour
- mandolin
- seen live
- bluegrass
- grateful dread

Station Tags:

- bluegrass
Meerkat

Spooky Lane  by Sam Bush
Walk On Boy  by Doc Watson
Mossy Cow  by Yonder Mountain String Band

'Spooky Lane' Tags:

- country influences
- theme time radio hour
- mandolin
- seen live
- bluegrass
- grateful dread

Station Tags:

- bluegrass
- mandolin

Add Tag
Meerkat

Spooky Lane by Sam Bush

Walk On Boy by Doc Watson

Mossy Cow by Yonder Mountain String Band

Down Where The River Bends by The Osborne Brothers

'Walk On Boy' Tags:
- raw music
- an upbeat two-step feel
- mandolin
- minor key tonality
- folk
- singer-songwriter

Station Tags:
- bluegrass
- mandolin
Music annotation

- The key behind keyword based system is annotating the music with tags
  - dance, instrumental, rock
  - blues, saxophone, cool vibe
  - pop, ray charles, deep

Ideas?
Annotating music

- The human approach

“expert musicologists” from Pandora

Pros/Cons?
Annotating music

- Another human approach: games
Annotating music

the web: music reviews

challenge?
Automatically annotating music

Learning a music tagger

song signal

review

tagger

blues, saxophone, cool vibe
Automatically annotating music

Learning a music tagger

What are the tasks we need to accomplish?
System Overview

Data
- Training Data
- Annotation

Features
- Vocabulary
- Document Vectors (y)
- Audio-Feature Extraction (X)

Model
- Parameter Estimation
Automatically annotating music

First step, extract “tags” from the reviews

**Frank Sinatra** - Fly me to the moon
This is a jazzy, singer / songwriter song that is calming and sad. It features acoustic guitar, piano, saxophone, a nice male vocal solo, and emotional, high-pitched vocals. It is a song with a light beat and a slow tempo.

**Dr. Dre** (feat. Snoop Dogg) - Nuthin' but a 'G' thang
This is a dance poppy, hip-hop song that is arousing and exciting. It features drum machine, backing vocals, male vocal, a nice acoustic guitar solo, and rapping, strong vocals. It is a song that is very danceable and with a heavy beat.
Automatically annotating music

First step, extract “tags” from the reviews

**Frank Sinatra** - Fly me to the moon
This is a jazzy, singer / songwriter song that is calming and sad. It features acoustic guitar, piano, saxophone, a nice male vocal solo, and emotional, high-pitched vocals. It is a song with a light beat and a slow tempo.

**Dr. Dre** (feat. Snoop Dogg) - Nuthin' but a 'G' thang
This is a dance poppy, hip-hop song that is arousing and exciting. It features drum machine, backing vocals, male vocal, a nice acoustic guitar solo, and rapping, strong vocals. It is a song that is very danceable and with a heavy beat.
Content-Based Autotagging

Learn a probabilistic model that captures a relationship between **audio content** and **tags**.

\[ p(\text{tag} \mid \text{song}) \]

- ‘Jazz’
- ‘Male Vocals’
- ‘Sad’
- ‘Slow Tempo’
Modeling a Song

Bag of MFCC vectors

cluster feature vectors
Modeling a **Tag**

1. Take all songs associated with tag \( t \)
2. Estimate ‘features clusters’ for each song
3. Combine these clusters into a single representative model for that tag

\[
p(x|t)
\]

Tag Model
1. Calculate the likelihood of the features for a tag model
Annotation

Semantic Multinomial for “Give it Away” by the Red Hot Chili Peppers
The CAL500 data set

The Computer Audition Lab 500-song (CAL500) data set

- 500 ‘Western Popular’ songs
- 174-word vocabulary
  - genre, emotion, usage, instrumentation, rhythm, pitch, vocal characteristics
- 3 or more annotations per song
- 55 paid undergrads annotate music for 120 hours

Other Techniques

1. Text-mining of web documents
2. ‘Human Computation’ Games - (e.g., Listen Game)
Retrieval

The top 3 results for - “pop, female vocals, tender”

1. Shakira - The One
2. Alicia Keys - Fallin’
3. Evanescence - My Immortal
## Retrieval

<table>
<thead>
<tr>
<th>Query</th>
<th>Retrieved Songs</th>
</tr>
</thead>
</table>
| ‘Tender’               | Crosby, Stills and Nash - Guinnevere  
                          Jewel - Enter from the East  
                          Art Tatum - Willow Weep for Me  
                          John Lennon - Imagine  
                          Tom Waits - Time |
| ‘Female Vocals’        | Alicia Keys - Fallin’  
                          Shakira - The One  
                          Christina Aguilera - Genie in a Bottle  
                          Junior Murvin - Police and Thieves  
                          Britney Spears - I'm a Slave 4 U |
| ‘Tender’ AND “Female Vocals” | Jewel - Enter from the East  
                              Evanescence - My Immortal  
                              Cowboy Junkies - Postcard Blues  
                              Everly Brothers - Take a Message to Mary  
                              Sheryl Crow - I Shall Believe |
Annotation results

Annotation of the CAL500 songs with 10 words from a vocabulary of 174 words.

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>0.14</td>
<td>0.06</td>
</tr>
<tr>
<td>Our System</td>
<td>0.27</td>
<td>0.16</td>
</tr>
<tr>
<td>Human</td>
<td>0.30</td>
<td>0.15</td>
</tr>
</tbody>
</table>
## Retrieval results

<table>
<thead>
<tr>
<th>Model</th>
<th>AROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>0.50</td>
</tr>
<tr>
<td>Our System - 1 Word</td>
<td>0.71</td>
</tr>
<tr>
<td>Our System - 2 Words</td>
<td>0.72</td>
</tr>
<tr>
<td>Our System - 3 Words</td>
<td>0.73</td>
</tr>
</tbody>
</table>