



Similarity by Metadata

Simon Putzke Mirco Semper
Freie Universität Berlin, Institut für Informatik

Seminar Music Information Retrieval

Motivation

Metadata

Tagging

Similarities

Webmining

Example

Motivation

Metadata

Tagging

Similarities

Webmining

Example

Why to compare by metadata?

- ▶ Metadata provides additional informations about songs
- ▶ Metadata can be emotionally different for people

What do we want to do?

- ▶ Use different methods to gain metadata for a song
- ▶ Compare songs for similarity on this metadata
- ▶ Create metrics for similarity

Motivation

Metadata

Tagging

Similarities

Webmining

Example

Types of Metadata in MIR

- ▶ Acoustic Metadata
- ▶ Cultural Metadata
- ▶ Editorial Metadata:
 - ▶ Title
 - ▶ Album name
 - ▶ Band/artst name
 - ▶ Genre
 - ▶ ...
- ▶ ... or the band
 - ▶ Band member
 - ▶ Instruments played
 - ▶ Genre
 - ▶ ...

Motivation

Metadata

Tagging

Similarities

Webmining

Example

- ▶ Automatic Tagging
 - ▶ Let the computer do the work
- ▶ Social Tagging
 - ▶ Let the community share the work
- ▶ Tagging as a Game
 - ▶ Have fun doing the work

- ▶ Webmine music from the Internet™
- ▶ Use witty algorithms to mine metadata given by single pieces of informations
 - ▶ Title leads to Artist
 - ▶ Artist leads to Genre
 - ▶ Artist leads to played Instruments
 - ▶ ...
- ▶ Tag the song with these informations

We follow up on this, later. Be patience.

- ▶ Having a community with a music database
- ▶ Each user tags his own music and shares the information
- ▶ Example: Last.fm

- ▶ Memory and Context
- ▶ Task Organization
- ▶ Social Signalling
- ▶ Social Contribution
- ▶ Play and Competition
- ▶ Opinion Expression

(Marlow et al., 2006; Ames & Naaman, 2007)

Distribution of the 500 most frequent tag types on Last.fm

Tag Type	Frequency	Examples
Genre	68%	heavy metal, punk
Locale	12%	French, Seattle, NYC
Mood	5%	chill, party
Opinion	4%	love, favourite
Instrumentation	4%	piano, female vocal
Style	3%	political, humor
Misc	2%	Coldplay, composers
Personal	1%	seen live, I own it
Organizational	1%	check out

(Source: Paul Lamere, Social Tagging and Music Information Retrieval, 2008)

- ▶ Play a game, tagging given music, earn points for tags
- ▶ More points for popular tags
- ▶ Example: MajorMiner

- ▶ Listen to a 10s Clip
- ▶ Earn points by describing the song
 - ▶ Only one other player tagged the same: 1 point
 - ▶ No other player tagged the same: 0 points
 - ▶ If at least two other tagged the same: 0 points
 - ▶ First tagger, but not the only one: 2 points

play me: <http://www.majorminer.org/>

- ▶ Train an algorithm to classify data with this data set
- ▶ Train the same algorithm with a dataset of Last.fm
- ▶ Comparison of both cases
 - ▶ MajorMiner: 67,2% accuracy
 - ▶ Last.fm: 62,6% accuracy

Motivation

Metadata

Tagging

Similarities

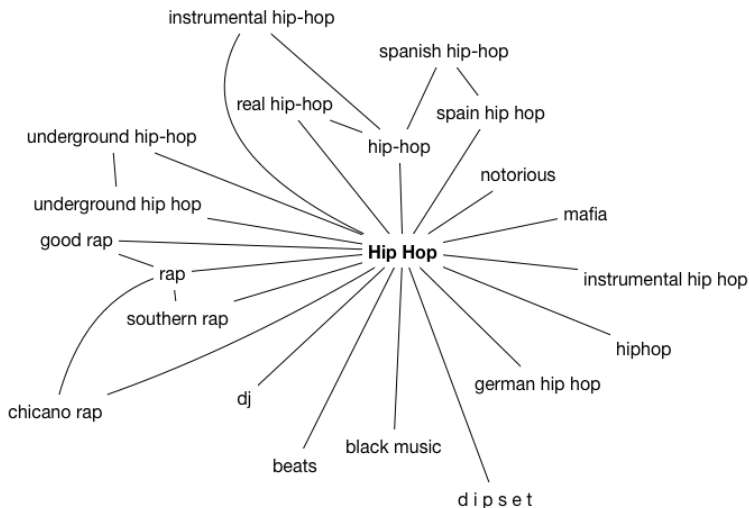
Webmining

Example

- ▶ Tag similarity and clustering
- ▶ Latent Semantic Analysis
- ▶ Item similarity

- ▶ Tags with same meaning may be spelled differently
 - ▶ Different spelling, eg. UK/US english
 - ▶ Spelling mistakes
 - ▶ Synonyms
- ▶ *Wrong* Tags are used
 - ▶ *Slow reggae* instead of *Rocksteady*
- ▶ Determine similar tags and cluster these into a cloud

One tag - many versions



(Source: Derived from 'Paul Lamere, Social Tagging and Music Information Retrieval, 2008')

- ▶ Tags are used to find content
- ▶ Misspelled tags are likely to be looked up
 - ▶ These are like synonyms
- ▶ Building up a cluster of misspelled tags
- ▶ Improve tagging search by tag similarities

Different spellings of one word

Top 22 *Female vocalists* variants with relative frequency

Tag	Freq	Tag	Freq
Female vocalists	100	Girl	< 1
Female	18	Women	< 1
Female vocalist	7	Chick	< 1
Female vocals	3	Ladies	< 1
Female voices	2	Femal voice	< 1
Female artists	2	Femaile voice	< 1
Female fronted	1	woman	< 1
Female vocal	1	Femaile vocalist	≪ 1
Female singers	1	girl music	≪ 1
Female fronted	1	girls with voices	≪ 1
Grrl	< 1	Girly girly	≪ 1

(Source: Paul Lamere, Social Tagging and Music Information Retrieval, 2008 (sic!))

- ▶ Tags are assumed to be a representation of a large pool of tags
- ▶ Assumption: there is a *latent* structure, relating items defining a *semantic space*
- ▶ Using statistical techniques, tag noise can be reduced
- ▶ A query with a tag can return items not associated with that tag
 - ▶ In the semantic space, objects are positioned relative to each other

- ▶ Often used tags are not very descriptive
- ▶ Use a weighting technique for more descriptive tags
- ▶ *Term frequency-inverse document frequency*
 - ▶ Determine the proportion of a tag in one document to all documents
 - ▶ Filter out the most used tags, concentrate on lesser used tags
- ▶ *Geek Rock* and *Nerd Rock* is more descriptive than *Rock*

- ▶ Popular vs unpopular songs
- ▶ Synonymy and noise
- ▶ Hack the tag

- ▶ Popular items are tagged more often
- ▶ New songs need to be tagged
- ▶ Too few tags cannot be used effectively
- ▶ Tagging games can help to solve these problems

- ▶ Unrelated tags can seem to be related.
 - ▶ 'Love': romantic song or 'I like it'
 - ▶ as seen before, can still be helpful
- ▶ Unuseful spam
 - ▶ 'Asdf'
 - ▶ 'Random'
 - ▶ 'i like pizza'

- ▶ Unpopular artists fake-tag for popularity
- ▶ Wrong tags to competitor
- ▶ Vandalism

Most frequent tags applied to Paris Hilton

Tag	Frequency
Brutal Death Metal	558
All things annoying in the world put together into one stupid b-tch	486
Crap	273
Officially Sh-t	238
Pop	231
Your ears will bleed	126
Emo	107
Whore	98
Best Singer in the world	72
Female Vocalist	60

(Source: Paul Lamere, Social Tagging and Music Information Retrieval, 2008 (sic!))

Motivation

Metadata

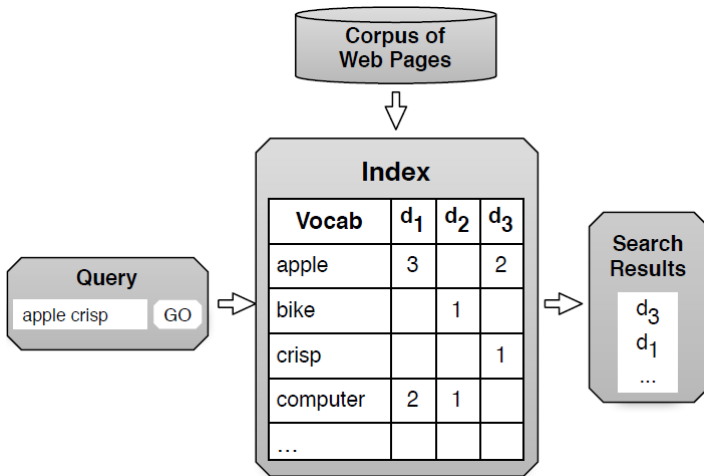
Tagging

Similarities

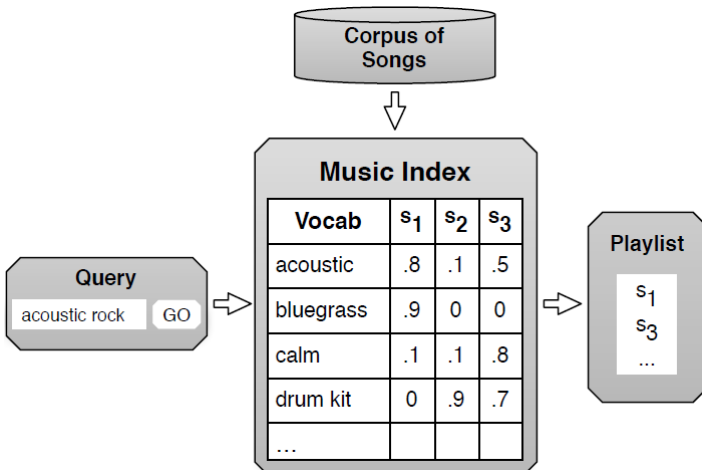
Webmining

Example

- ▶ Incorporation of semantics
- ▶ Instant availability
- ▶ Time-awareness



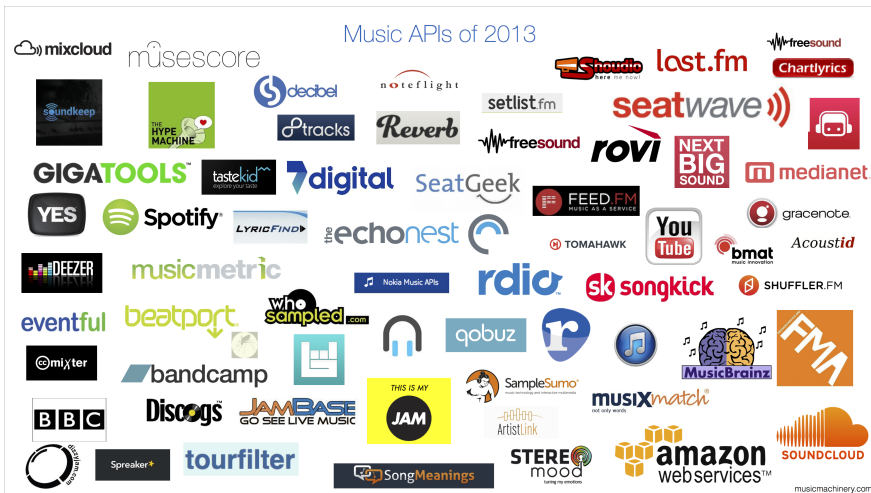
(Source: Douglas Turnbull, Indexing Music with Tags, 2012)



(Source: Douglas Turnbull, Indexing Music with Tags, 2012)

- ▶ Conducting a survey?
- ▶ **Harvesting social tags?**
- ▶ Playing annotation games?
- ▶ **Mining web documents**
- ▶ Autotagging audio content

Harvesting social tags



(Source: <https://musicmachinery.com/music-apis/>, checked at 13.11.2016)

- ▶ Scrobbling
- ▶ Tags
- ▶ Plugins for iTunes, Winamp etc.
- ▶ More focused on the data lately

- ▶ Cold-start for new songs
- ▶ Popularity bias
- ▶ Tags centered on artists
- ▶ Malicious behaviour

- ▶ Query search engines
- ▶ Domain focused crawlers
- ▶ Example Queries:
 - ▶ "<artist name>" music
 - ▶ "<artist name>" "<song name>" music review
 - ▶ Suffix: "<site url>" for restriction of sites

- ▶ Noisy
- ▶ Short-head queries better than long-tail
- ▶ Data is often artist focused

Motivation

Metadata

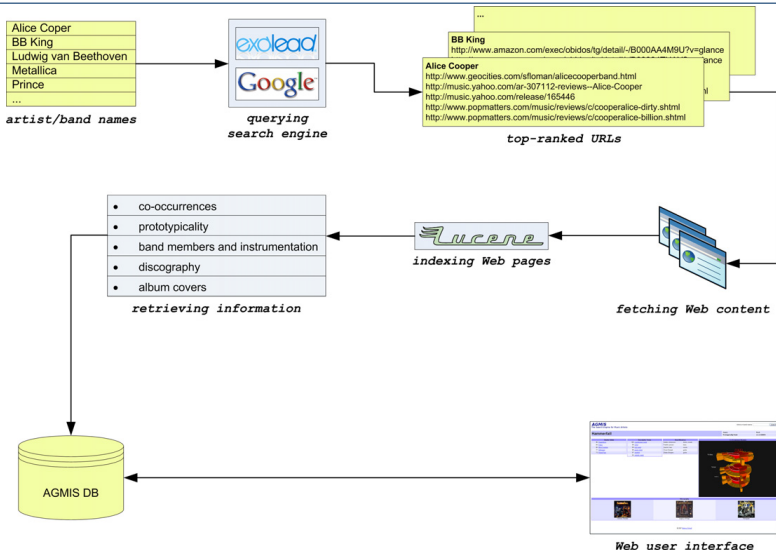
Tagging

Similarities

Webmining

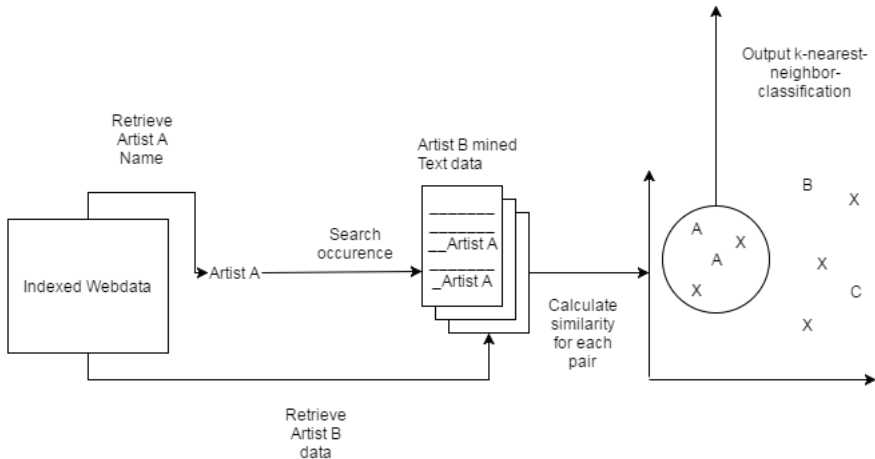
Example

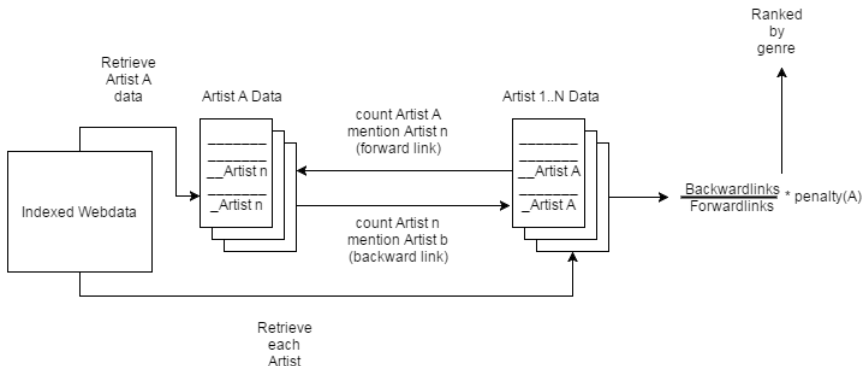
A MIS automatically generated with Web Mining



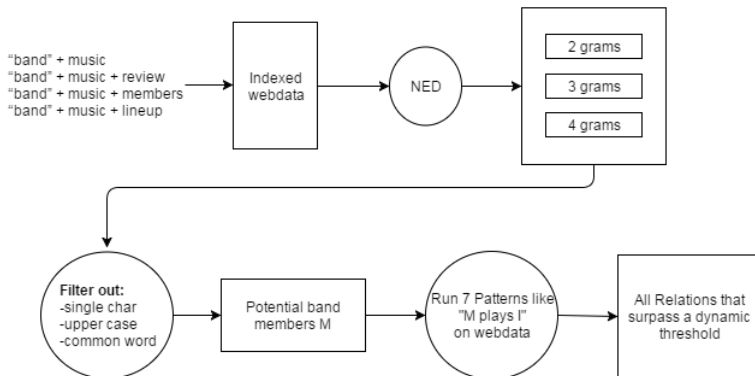
(Source: Markus Schedl, A music information system automatically generated via Web content mining techniques, 2010)

Similarity of artists and bands

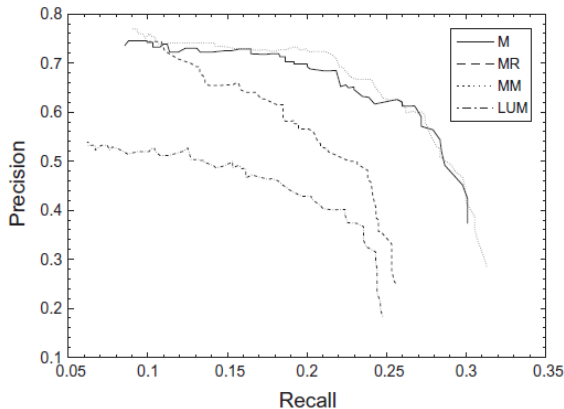




Find band members and instrumentation



Find band members and instrumentation results



"band" + music M

"band" + music + review MR

"band" + music + members MM

"band" + music + lineup LUM

(Source: Markus Schedl, A music information system automatically generated via Web content mining techniques, 2010)

- ▶ Using 1500 music related term list
- ▶ Weighted via DF, TF and TF-IDF
- ▶ Scores of 2.22, 2.43, and 1.53 for TF, DF, and TF-IDF

Thanks for listening

Thank you for your
attention.
Questions?

- ▶ Markus Schedl, A music information system automatically generated via Web content mining techniques, 2010
- ▶ Douglas Turnbull, Indexing Music with Tags, 2012
- ▶ Michael I. Mandel & Daniel P.W. Ellis, A Web-Based Game for Collecting Music Metadata, 2008
- ▶ Scott Deerwester, Indexing by Latent Semantic Analysis, 1990
- ▶ François Pachet, Knowledge Management and Musical Metadata, 2005
- ▶ Stephan Baumann, Using cultural metadata for artist recommendations, 2003
- ▶ William W. Cohen, Web-collaborative filtering: recommending music by crawling the Web, 2000
- ▶ Paul Lamere, Social Tagging and Music Information Retrieval, 2008

- ▶ <https://en.wikipedia.org/wiki/Last.fm>
- ▶ <https://www.washingtonpost.com/news/arts-and-entertainment/wp/2014/03/27/the-slow-sad-death-of-last-fm/>