

Automatically Extracting, Analyzing, and Visualizing Information on Music Artists from the Web

Workshop on Learning the Semantics of Audio Signals (LSAS)
21st June 2008, Paris, France



Markus Schedl, Peter Knees

Department of Computational Perception
Johannes Kepler University Linz

markus.schedl@jku.at
<http://www.cp.jku.at>

Overview

- Introduction
- Motivation for an Automatically Generated Music Information System
- Data Processing Pipeline for Web Information Retrieval
- Information Extraction
 - Artist Similarity
 - Prototypicality of Artist for a Genre
 - Album Cover Artwork
 - Band Members and Instrumentation
 - Descriptive Terms (Tagging & Visualization via Co-Occurrence Browser)
- Future Work

Example of a Music Information System: *last.fm*

last.fm the social music revolution

Music Users Listen **NEW!** Events Charts Tools

Sign Up and create a profile

You are not logged in | Log in | Help | Music Search

MC Solaar

Overview | Pics | Bio | Events | Albums | Listeners | Similar | Charts | Tags | Journals

Meizu 8GB MiniPlayer Taking pre-orders. 4GB only \$126 Your discount miniPlayer source

Radio Mp3 Easily Search The Web For The Top Deals On Radio Mp3 Here!

Ads by Google

Not what you wanted? [View full search results...](#)

MC Solaar (read more)

399,909 plays scrobbled on Last.fm

MC Solaar is the stage name of francophone [hip hop](#) artist Claude M'Barali (born March 5, 1969). By far the most internationally popular French rapper, he was born in Dakar, Senegal to parents from Chad. The family moved to Villeneuve-Saint-Georges, just outside Paris, when Solaar was six months old. He also spent time at th... [\(read more\)](#)

[Edit this artist description](#)

Listen Now

MC Solaar - Nouveau Western

Track	Time	Listeners
Nouveau Western	0:30	3,476
Caroline	0:30	3,277
Gangster Moderne	0:30	2,218
Les Temps Changent	0:30	2,363

MC Solaar's Radio (show all stations)

Play MC Solaar's **Similar Artists** [Play in pop up](#)

Play MC Solaar's **Top Listeners** [Play in pop up](#)

Top Albums (see all)

Similar Artists

- IAM
- The Roots
- Guru
- Saïan Supa Crew
- De La Soul
- Outkast
- Fugees
- DJ Shadow

User Tags (see more)

francais french french hip hop french hip-hop french rap hip hop hip-hop rap

[Tag this artist](#)

Weekly Top Listeners (see more)

29,740 listeners total

- teh126
- suesand
- bibu
- boltzy
- 1Mike1
- yoshiimil

Related Journals (read more)

- Top 25 thingy** by [Miltec](#)
- die ersten wochencharts.** by [preemoe](#)
- ...L' Attente (New Albums I'm Waiting 4)** by [babyboy75](#), 1 comment
- O Crêpe! playlist - le 20 mars 2007** by [choupinette](#), 1 comment

[Write a journal about this artist!](#)

Related Groups (see all)

- [Ethnicity in hip hop](#)
- [Europeans Against The World](#)
- [Jazz Hop](#)
- [DJ Premier](#)
- [The Music Chain](#)
- [Hip-Hop, Noise, Experimental, and anything Crazy](#)

Example of a Music Information System: *allmusic*

allmusic
allmusic allmovie allgame

GO
Artist/Group advanced search

Celebrating 15 years AMG All Media Guide

» Classical Corner » Artist Spotlight
» Top Composers » Classical Reviews

You are not logged in.
Login or Register

ROCK JAZZ R&B RAP COUNTRY BLUES WORLD ELECTRONICA CLASSICAL MORE...

Overview Biography Discography Songs Credits Charts & Awards

Dragonforce
wrong person? more matches HERE

Send to Friend

Biography by James Christopher Monger
U.K. power metal sextet Dragonforce formed in 1999 around the twin-guitar assault of Herman Li and Sam Totman, keyboard player Vadim Pruzhanov, and the powerful vocals of ZP Theart -- drummer David Mackintosh and bass player Adrian Lambert joined later. Dragonforce had already toured with the likes of Halford and Stratovarius before the release of their debut, 2003's *Valley of the Damned*. The success of that record found the melodic rockers a worldwide audience, resulting in a string of sold-out concerts in Asia and Europe. The group's sophomore effort, *Sonic Firestorm*, was released in 2004, followed by *Inhuman Rampage* in 2006.

Picture Browser
< Previous Next >

Formed
1999 in London, England

Years Active
1910 20 30 40 50 60 70 80 90 2000

Genre Styles
Rock Power Metal Neo-Classical Metal Heavy Metal

Moods
Wistful Indulgent Dramatic Theatrical Exuberant Aggressive Wintry Playful Passionate Fiery

AMG Artist ID
P 554753

Corrections to this Entry?

Think You've Seen It All?
Check Out The Latest Viral Videos

Viral Videos:
Editor's Picks
JibJab
StupidVideos

★ Editor's Picks

msn Video Watch the Latest!

buy listen buy listen buy listen

Watch music videos by this artist!

Group Members
Herman Li
Sam Totman
ZP Theart
Vadim Pruzhanov
Adrian Lambert
David Mackintosh

Influenced By
Journey

Performed Songs By
Sam Totman
ZP Theart
Vadim Pruzhanov
Herman Li

Similar Artists
Helloween
Manstead

The Big Picture

- Creating an automatically generated/populated music information system (AGMIS)
- How ?
 - Web Content Mining (Text, Image, Audio, Video)
- Using techniques from Information Retrieval (IR) and Natural Language Processing (NLP)

Motivation for AGMIS

- No need for labor-intensive maintenance of the system (no music experts, nor large community needed)
- Not vulnerable to editors' cultural bias (allmusic), nor to vandalism (last.fm)
- Automatical incorporation of new information as soon as they become available on the Web

What AGMIS Will Look Like

AGMIS

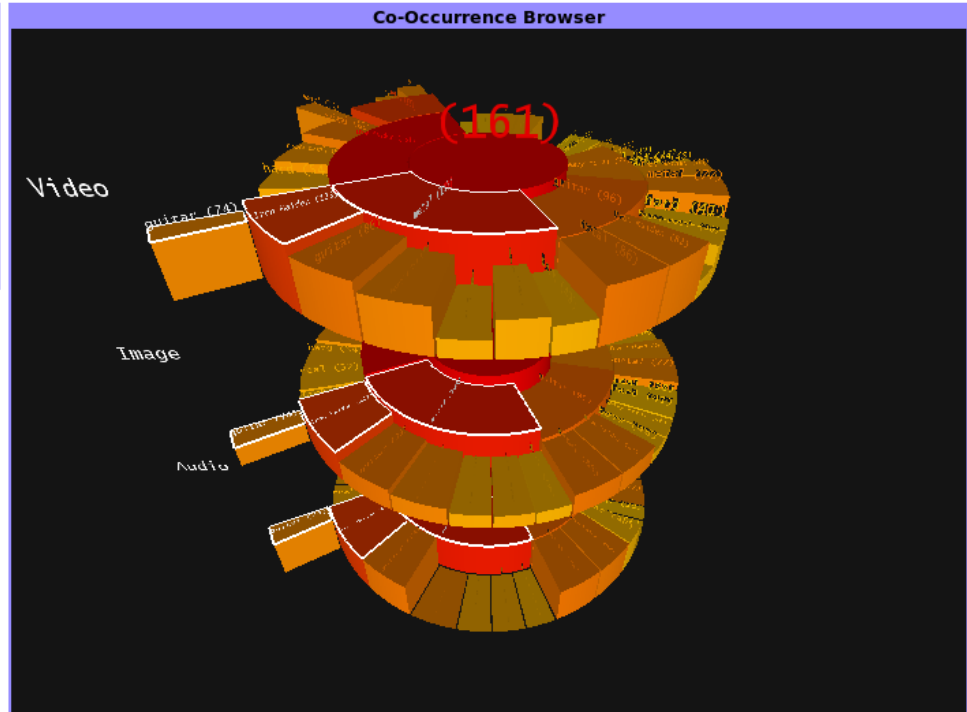
The Search Engine for Music Artists

Artist or band name:

Hammerfall

Genre:	Rock
Prototypicality Rank:	n/a of 268878

Similar Artists	Descriptive Terms	Band Members
96 Dragonforce	100 scandinavian metal	Anders Johansson drums, vocals
84 Edguy	99 metal	Fredrik Larsson bass
80 Blind Guardian	99 true metal	Joacim Cans vocals
77 Helloween	98 power metal	Oscar Dronjak guitar
73 Primal Fear	97 swedish	Stefan Elmgren guitar
	84 melodic metal	

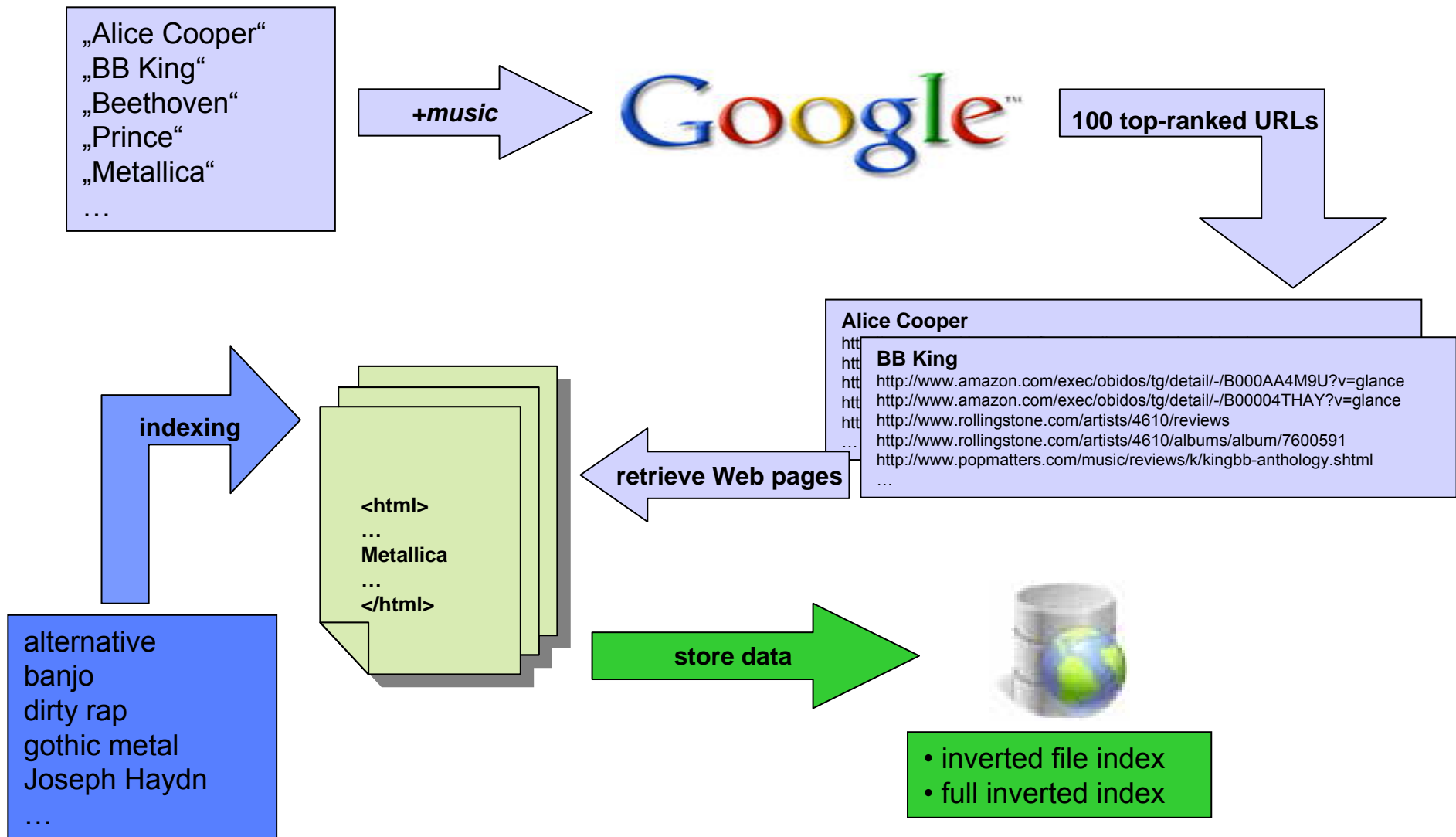


Discography		
		
Crimson Thunder	Legacy of Kings	Renegade

Parts of AGMIS

- Similar and prototypical artist detection
- Album cover retrieval
- Band member and instrumentation detection
- Automatic attribution/tagging of artists
- UI to browse artist-related Web pages (Co-Occurrence Browser)

Data Processing Pipeline



Similar and Prototypical Artist Detection

- Data source: inverted file index
- Calculate document frequency (DF) of artist name v on Web pages retrieved for artist u
- Estimate conditional probability for artist v to be found on an arbitrary Web page of u (relative frequency DF_{uv} / DF_{uu})
→ asymmetric conditional probabilities
- Compute arithmetic mean to derive a symmetric **artist similarity measure**
- Use asymmetric probabilities to estimate **prototypicality of an artist for a genre** (idea: within a genre, Web pages about less prototypical artists tend to mention more prototypical artists more frequently than vice versa)

Similar and Prototypical Artist Detection: Evaluation

- Artist similarity:

On collection of 224 well known artists from 16 general genres (Rock, Classical, Blues, ...):

classification accuracy (k-NN, leave-one-out CV) of about 85%

On collection of 103 artists grouped in 22 quite specific genres (Bossa Nova, Death Metal, Jazz Guitar, German Hip-Hop, ...):

classification accuracy (k-NN, leave-one-out CV) up to 70%

- Artist prototypicality:

On collection of 1,995 artists from 9 genres:

overall agreement with importance ranking by AMG: 60-65%

Album Cover Retrieval

- Data source: full inverted index (word level + HTML tags)
- Image Pre-Filtering (quadratic, scanned CDs)
- Different approaches for image selection:
 - char/tag distance of artist and album names to tag, select image with lowest distance
 - calculate an average histogram, select image which is nearest to it
 - use the first image returned by Google's image search (baseline)

Album Cover Retrieval: Evaluation

- Test set: 3,311 album names
- Best results using pre-filtering (quadratic constraint and scanned compact disc filter):

<i>approach</i>	<i>correct</i>
Google's image search (baseline)	56.7%
Avg. Histogram	10.0%
Tag distance	58.9%
Char distance	57.9%

Band Members and Instrumentation

- Data source: full inverted index
- Named Entity Detection to find candidate members (N-grams of capitalized words, filtering of common speech words)
- Rule-based Linguistic Analysis
 1. M plays the I
 2. M who plays the I
 3. R M
 4. M is the R
 5. M, the R
 6. M I
 7. M R

M: member, I: instrument,

R: role (singer, guitarist, bassist, drummer, keyboardist)

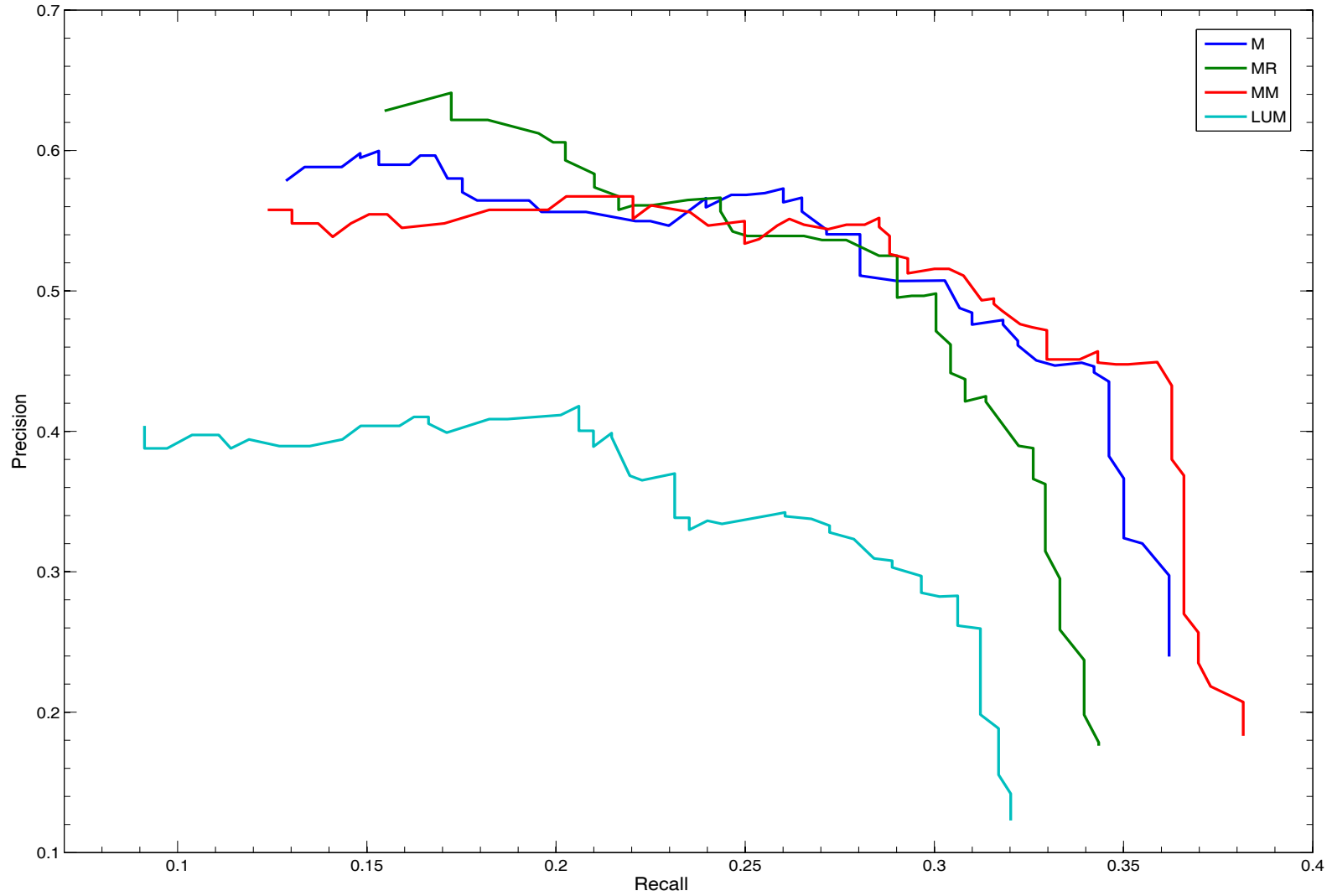
Band Members and Instrumentation (2)

- Calculate number of rule appliance
→ (member, instrument, rule, DF)
- Combine information over all rules
→ (member, instrument, $\sum DF$)
- Discard uncertain information, i.e., (member, instrument)-pairs with $\sum DF$ value below a threshold t_{DF}
- Predict remaining (member, instrument)-pairs
→ m:n assignment between member and instrument

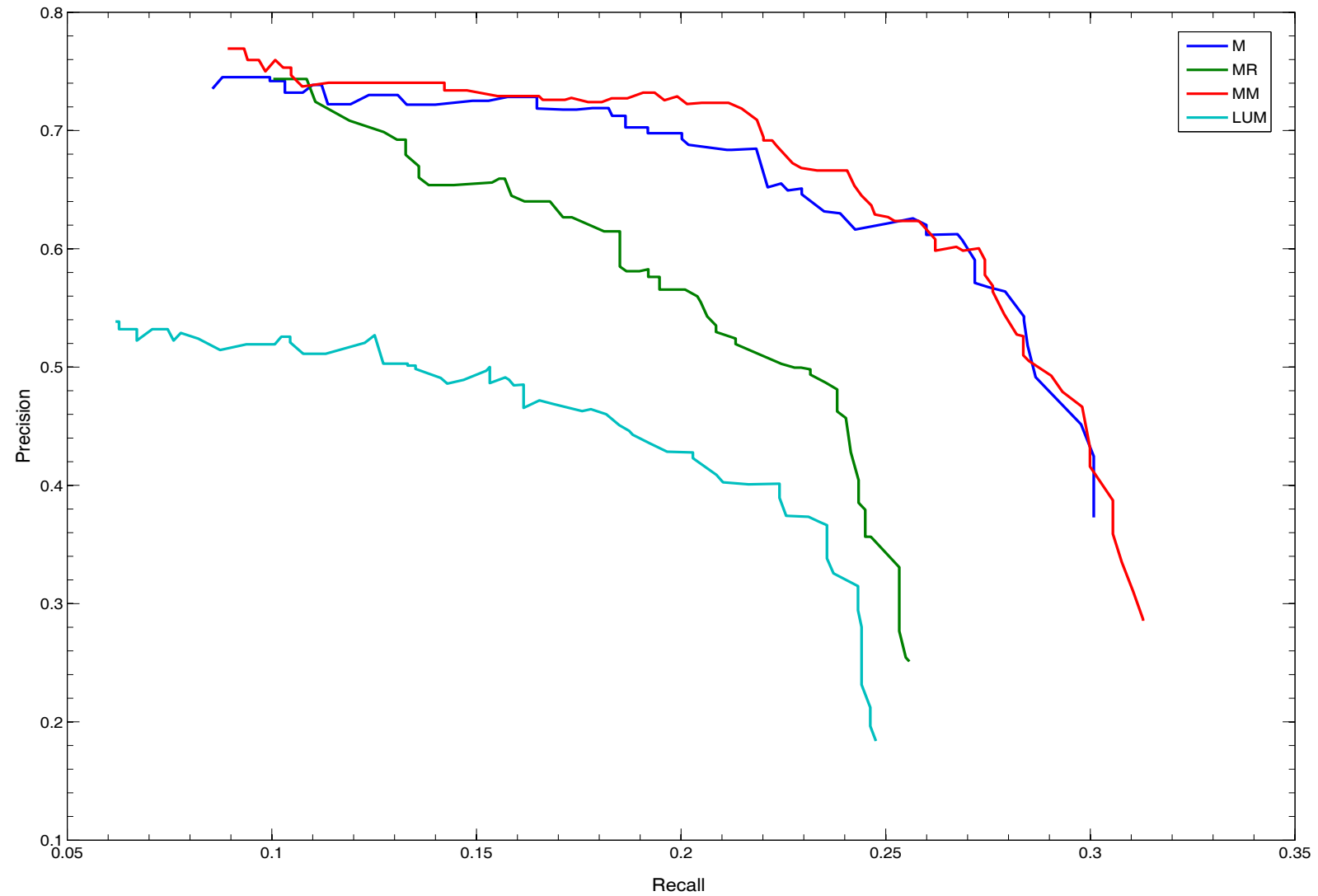
Band Members and Instrumentation: Evaluation

- 2 ground truth sets containing line-up of 51 bands
 - M_c 240 current band members
 - M_f 499 current and former members
- Measure precision and recall (set of predicted band members vs. set of band members given by ground truth)
- Upper limit for achievable recall: about 60%

Band Members and Instr.: Results M_C



Band Members and Instr.: Results M_f



Automatic Attribution/Tagging of Artists

- Data source: full inverted file index
- Different term weighting functions (TF, DF, TFxIDF) to rank terms from music dictionary occurring on corpus of artist's web pages
- User study to assess descriptiveness of highest ranked terms, using the three different weighting functions:
 - 112 well known artists from 14 genres
 - Web page indexing using dictionary of 1,506 musically relevant terms
 - 10 highest ranked terms of the 3 weighting functions merged
 - 1 term set for each artist
 - 5 participants, each told to rate terms for the artists they knew well (categorizing each term in three classes: +, -, ~)

Automatic Attribution/Tagging of Artists: Results

- 172 individual artist ratings returned
- 92 of 112 artists covered
- Overall excess of good terms (+) over bad terms (-)
 - TF: 2.22
 - DF: 2.42
 - TFxIDF: 1.53
- TF and DF performed significantly better than TFxIDF, no significant difference between TF and DF

Browsing Artist-Related Web Pages via COB

- Data source: full inverted file index
- Create co-occurrence tree and visualize it

Algorithmic outline:

1. Start at root node (set of all Web pages retrieved for artist)
2. Select i most important terms t_i (according to term weighting func.)
3. Create new sets of Web pages containing terms: $\{t_1\}, \dots, \{t_i\}$
4. For each of these sets, goto 2. (until maximum depth reached)

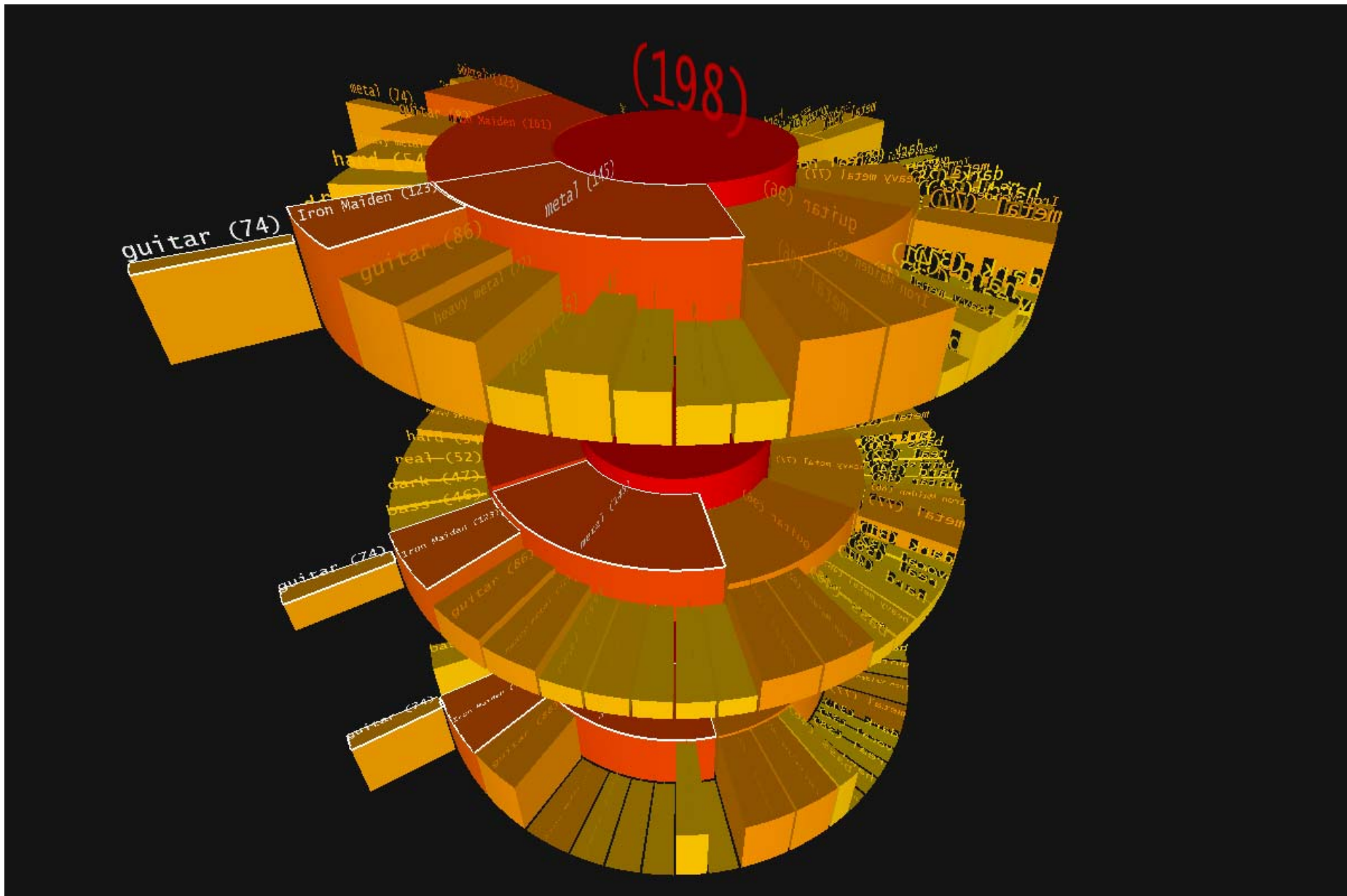
Sunburst / InterRing Visualization

- Circular, space-filling visualization technique
- Center represents root node
- Deeper elements in hierarchy are drawn further away from center
- Children are drawn within angular borders of their parent

Co-Occurrence Browser

- Brings the Sunburst to 3D
- Additional data dimension can be encoded in height of each arc
- Stacking a number of such 3D-Sunbursts offers even more dimensions
- Arcs illustrate Web pages with certain term combinations (the most important ones according to some term weighting function)
- Amount of multimedia data found on the artist-related Web pages is visualized (three layers – for audio, image, and video files)
- User interaction by rotating, zooming, changing the view angle, displaying Web pages and multimedia content

Co-Occurrence Browser



Future Work

- Combine with audio signal-based MIR (incorporate real music)
- Automatically detect country of origin for artists
- Automatic biography generation
- Automatically detect new artists on the Web