jWebMiner
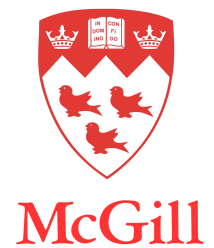
Gabriel Vigliensoni

Music Technology Area
Schulich School of Music
McGill University

Outline

Context

What Is It?

Where?

How?

Future

Context

Enormous (and growing exp) amount of information

Web : valuable source of information for
cultural features

Cultural Metadata / Community Metadata -> role in
how we organize and interpret music

How to access and retrieve it? -> Web Services

jMIR (LL, HL and CM)

What?

Software package to extract cultural features
from the web

Through Web Services of Yahoo! and Google

Accesses search engines and acquires
hit counts from search strings

Co-occurrence / Cross-tabulation

Flexible interface / Extensible API

Where?

Artist similarity / recommendation

Genre classification

Improve MIR research

How?

Analyze iTunes XML, ACE XML,
Weka ARFF or text string

Accesses web to measure
co-ocurrence or cross-validation

Offers variety of metrics and scoring systems

Options : string synonims, string filtering,
site weighting and limiting

Outputs ACE XML, Weka ARFF or text files

String Analysis
Co-Ocurrence Extraction

String Analysis
Co-Ocurrence Extraction

String Analysis
Cross Tabulation Extraction

String Analysis
Cross Tabulation Extraction

jWebMiner

Search Information

Search Words | Required Filter Words | Excluded Filter Words | Site Weightings | Options | Results

Save   SAVE FORMAT: HTML

Final feature scores:

	classical	grunge	jazz	pop	punk	reggae	RnB <SYNONIM> R&B	rock	soul
Bob Dylan	0.12363183317603281	0.10335031753993108	0.15014396358345858	0.11609743184389013	0.12128306759418533	0.09786167437565628	0.017616718527834408	0.12622557226281575	0.14378942109619566
Bruce Springsteen	0.115550675925774406	0.08729766615727598	0.14155112911738713	0.12992943395767312	0.11967640218813018	0.1075144726003357	0.02060717050909051	0.1393733903158024	0.13854357589656102
Elvis Presley	0.12347572224412023	0.06510940616824744	0.12654820066434086	0.11741713899185596	0.12924197741911173	0.16852974994385497	0.023673177038320267	0.11952176669330104	0.1264826083684762
Jimi Hendrix	0.09733116346702216	0.15194886047542291	0.13419817857769895	0.10177915082108702	0.1289661125081192	0.1095985455737742	0.027233241031936844	0.1280719503218512	0.12087279714039469
Marvin Gaye	0.08730661889529095	0.0565109545864871 8	0.13318096233730978	0.0903886727886739	0.10460704735633405	0.17362686374359915	0.07385617903102522	0.08574320496727018	0.19477949629400967
Michael Jackson	0.10065941290151373	0.055627748980125555	0.08018465503806073	0.12629306169268686	0.0843347561152314	0.10703893882118404	0.2596092814681184	0.09601636294642804	0.090235782036 6512
Miles Davis	0.13926397018459116	0.05404603793151556	0.33314023529147385	0.07981565417197901	0.07788413018555447	0.0935291194589407	0.017041445143143475	0.08714237006060957	0.1181370375721 9228
Nirvana	0.05453541396709075	0.388958259283461	0.06940944743686153	0.07892246449974273	0.13663657798306628	0.08391995827331947	0.010169575088799697	0.08679334684445061	0.09065495662320794
The Beach Boys	0.1307907641272894	0.10148957288678868	0.1325529769360459	0.1240361988274823 7	0.12179101179654514	0.12297594249395619	0.011963638246183524	0.11834329970069034	0.13605659498501865
The Beatles	0.16592234553484114	0.11070958526068023	0.12084995529331216	0.09897173939704919	0.11200207753975672	0.10014478103816198	0.014793665355849885	0.15468310109216596	0.1219227494881 8278
The Clash	0.08660505696839255	0.1384604308696394	0.08541848726406848	0.0948736588560785	0.2376437367196339	0.12718646105118395	0.008938049350127309	0.10625950122542085	0.11461461769545496
The Rolling Stones	0.12004934807228002	0.10294625271070068	0.1218493233480143 6	0.09921329430923036	0.1508254090310163	0.1650785788218743 2	0.011584010148312979	0.10530751125607528	0.12314714043041053
The Velvet Underground	0.07778774661342816	0.15498317061961958	0.09481883399413164	0.10991161566484975	0.21103723789538667	0.12388842817774205	0.01171084260693621 6	0.10173481771183789	0.11412730671606806
Van Morrison	0.12732197749390156	0.07956462069641826	0.15412817095561257	0.11401903658368119	0.1184476206300117 5	0.1343454727336707	0.00632683181230282 2	0.11263362859021121	0.15312356596449353

String Filtering
Required Filter Words

String Filtering
Required Filter Words : 'music'

jWebMiner

Search   Information

Search Words | Required Filter Words | Excluded Filter Words | Site Weightings | Options | Results

Save   SAVE FORMAT: HTML

Final feature scores:

	classical	grunge	jazz	pop	punk	reggae	RnB <SYNONIM> R&B	rock	soul
Bob Dylan	0.11359070679767803	0.11127127069169995	0.14640325772602641	0.11869663225760353	0.12360284624104792	0.09391284380631182	0.019529889091934788	0.1326242350179311	0.14036831836976643
Bruce Springsteen	0.03802079743342982	0.10101912031923468	0.1440803687066064	0.14613417086851446	0.12915391100976062	0.1123470048411158	0.027461461350002006	0.15490606115799332	0.14654943614928867
Elvis Presley	0.12008579168355972	0.06424161048965861	0.13534877401579634	0.12090465217409022	0.13450572963121338	0.15731990150599587	0.025622941473060337	0.11283423071693703	0.1291363830968866
Jimi Hendrix	0.10659446971770008	0.15577048732568297	0.13057798347852484	0.10761694304035302	0.1239914797276787	0.09759440282116592	0.03181490369796203	0.1288096529657405	0.11722967722519194
Marvin Gaye	0.09379731601828739	0.0654302628575439	0.1353763014689327	0.1005231380806996	0.08734153198020474	0.15490885471440177	0.08578611340782488	0.08994150350404721	0.1868949779680579
Michael Jackson	0.10766066682491426	0.05903829405888059	0.0744636654941283	0.11702122679629087	0.08616982031039325	0.11061302002873172	0.26018360153365494	0.09614833328717004	0.08870137166583604
Miles Davis	0.14639278593095187	0.05898534062579828	0.3196328934870799	0.0879860195362154	0.07081165778553478	0.08338090933327887	0.02005010700923041	0.09265166522381868	0.12010862106809192
Nirvana	0.05845295558547615	0.3738361049808724	0.07152392172790223	0.08420160036946231	0.1368526817029831	0.08473370567512822	0.01149400856330067 2	0.08761208289378666	0.09129293850108824
The Beach Boys	0.1287286192499989	0.10845667221333499	0.127206793267316	0.12906285854114938	0.11588640729326982	0.128084975864807	0.012656398955205258	0.11996724748335223	0.1299500271315664
The Beatles	0.16163124911335597	0.10929506240772656	0.1304615197397995	0.10315855634466428	0.11072110123149637	0.10586654440315182	0.015370809274184393	0.1393798676049717	0.1241152898806493 8
The Clash	0.07115771086020688	0.15220821736211565	0.09146422095827084	0.1031848601896140 1	0.23980937079333425	0.10446185234683257	0.010513289358129316	0.11379577836906574	0.1134046997624308 4
The Rolling Stones	0.1163644427768892	0.10399701909501001	0.1212875810383473 7	0.1050124103704502 7	0.1424440175692711 6	0.1678568233004153 5	0.01263994665348018 4	0.10791527069918443	0.1224824884969519 4
The Velvet Underground	0.09679037953702123	0.05446434221372440 5	0.11531168521096478	0.1340600555609053	0.2323960617295896 7	0.09606011942356023	0.01723963871763232	0.1294445154531972	0.12423326606128295
Van Morrison	0.13540685496352833	0.0906180942150860 8	0.1522066480242895 4	0.11995401848710877	0.1068235528941743 5	0.1188792532806872 7	0.00697347240463834 1	0.11751377938315487	0.1516243263473324 7

String Filtering
Site Weighting : only www.allmusic.com

Search   Information

Search Words | Required Filter Words | Excluded Filter Words | Site Weightings | Options | Results

Save   SAVE FORMAT: HTML

Final feature scores:

	classical	grunge	jazz	pop	punk	reggae	RnB <SYNONIM> R&B	rock	soul
Bob Dylan	0.1439832205362147	0.050425071357143504	0.10810646452028846	0.149918994999024	0.10961763730655044	*0.1728266702011444*	0.0	0.15024977053366367	0.11487217054597078
Bruce Springsteen	0.12811235035168164	0.031451411062523235	0.10746929124307134	0.14694165286035865	0.15418768788079343	*0.186014215994374*	0.0	0.14158498045576962	0.10423841015142789
Elvis Presley	0.1910300023346305	0.008932878222692755	0.10990247947662975	0.16540401837377017	0.0512197657616845	0.159176957888847	0.0	*0.19352594929401346*	0.12080794864773184
Jimi Hendrix	0.12604840928452327	0.10458769376619535	0.1506531569285147	0.12725660978363332	0.0994280596867375	0.10873563422804826	0.0	0.1322373139572585	*0.1510531223650889*
Marvin Gaye	0.12065196834403161	0.008977143281814379	0.10585368697814249	0.1084171532855611	0.0381859101878223	0.1341030045801901	0.0	0.08329455490659603	*0.4005165784358420*
Michael Jackson	0.12708244689896248	0.08602599207529585	0.105807889507564	0.12745980767048182	0.046066225085050906	*0.2048397613214043*	0.0	0.11030974971542544	0.19240812772581511
Miles Davis	0.12830396679011974	0.0	*0.4686586232530201*	0.05154060281799454	0.031302215251731554	0.1311089032069541	0.0	0.05398143304952345	0.1351042556306565
Nirvana	0.046243063110970886	*0.6550636336659034*	0.024520251578243384	0.05669886761920044	0.11057691533866543	0.029533757982552464	0.0	0.05249417171193495	0.024869338992528817
The Beach Boys	0.1519200854369951	0.05487253812630972	0.08764327996039067	*0.1822206827892765*	0.10949557098933727	0.12881013800020855	0.0	0.17268527060970149	0.1123524340877808
The Beatles	0.1475802719515931	0.08880141892257161	0.08405774833933945	0.16146791496725021	0.1428588247521283	0.09375543848745284	0.0	*0.1670527879354533*	0.11442559464421119
The Clash	0.07874313890497822	0.08336964109530594	0.029644316411076163	0.09067058746816747	*0.373796994136578*	0.21349685867792628	0.0	0.09233906455107678	0.0379393987548907
The Rolling Stones	0.12813167576443674	0.07379902340273287	0.08513678565975755	0.14893203550691148	*0.1594541481200475*	0.11950710201195738	0.0	0.14946346134614957	0.1355757681880069
The Velvet Underground	0.10178729388703987	0.14868526990464512	0.0461956063298789	0.13255362049530248	*0.353335327398102*	0.05107758271856566	0.0	0.12591642511818624	0.0404488741482796
Van Morrison	0.12550437842713016	0.021879767831814702	0.15038542772561353	0.1411826250270223	0.07546181209452481	0.1564191162367917	0.0	0.13454316120982607	*0.194623711447276*

String Filtering
Site Weighting : 3 music sites

jWebMiner

Search   Information

Search Words | Required Filter Words | Excluded Filter Words | Site Weightings | Options | Results

Save   SAVE FORMAT: HTML

Final feature scores:

	classical	grunge	jazz	pop	punk	reggae	RnB <SYNONIM> R&B	rock	soul
Bob Dylan	0.10904959779365359	0.14546325087680828	0.09966543694307924	0.1230964767727568	0.12685981420705494	0.1519540884307979	0.0	0.13827993050636553	0.10563140446948346
Bruce Springsteen	0.0695726592173311	0.12124997225012582	0.11299018016075986	0.14705209211557363	0.2379739116854887	0.07242260838925978	0.0	0.07242527662090488	0.1663132995605601
Elvis Presley	0.3615402796912261	0.017706539596652761	0.1161648197122081	0.15860690474882577	0.03014875501781079	0.12452506771946409	0.0	0.11727478554669793	0.07467399159723981
Jimi Hendrix	0.2241943873183698	0.07082783305502148	0.18400417623664425	0.1395659507984872	0.07147505828869608	0.10876130063097438	0.0	0.09335584841580807	0.1078154452559987
Marvin Gaye	0.2275978873744530	0.09690333554955695	0.10908454574771584	0.1123767818741868	0.042684794356088164	0.1313935405957715	0.0	0.055431725015171913	0.22452738948705583
Michael Jackson	0.06160525095801007	0.10654434191492408	0.11981169737050705	0.1531567202066544	0.1654018596009209	0.15165747435666183	0.0	0.07183375482763951	0.169988900764512
Miles Davis	0.18153410851821924	0.0030924094706673588	0.38789373982043757	0.06342335197262297	0.06344504446610859	0.1799091837631163	0.0	0.03541638229390829	0.08528577969491985
Nirvana	0.0433541426343666	0.2726662861711876	0.08496155809050593	0.1282932493252948	0.18979989975569905	0.07519216650628371	0.0	0.05501338903114734	0.15071930848551493
The Beach Boys	0.3142492897637153	0.0491454293066514	0.10419098623266572	0.18518482070465167	0.046955945815668615	0.12202797377757496	0.0	0.10126877465305567	0.07702114053336032
The Beatles	0.1124487627997325	0.06309138883019193	0.10868617884595337	0.10252281043249001	0.060626140056580215	0.12838702065824922	0.0	0.3125811066502067	0.11165663187473576
The Clash	0.1211211319849531	0.1952007009369298	0.06991608215378967	0.09385401659964539	0.1499991256619293	0.2660937804266079	0.0	0.04759502796794906	0.05622013426819597
The Rolling Stones	0.2160787948533601	0.08276769602804901	0.11174279237491594	0.15811610291582218	0.11464946023402046	0.0987338287570128	0.0	0.09423057736978055	0.12368074746703893
The Velvet Underground	0.1696174610281419	0.1572578899168113	0.0835038675690353	0.16566250763386864	0.1927706877451766	0.041798430023199495	0.0	0.08368610295619427	0.1057030531757247
Van Morrison	0.2161443441218872	0.04061587904372937	0.18323850311672285	0.12907107387401778	0.045902299725245184	0.1678743943303142	0.0	0.08615075341286729	0.1310027523124988

String Filtering
Pattern-based filter string

String Filtering
Pattern-based filter string

Final feature scores:

	classical	grunge	jazz	pop	punk	reggae	RnB <SYNONIM> R&B	rock	soul
Bob Dylan	0.0	0.7022638033080241	0.053562493472645904	0.0	0.0	0.03807454355284468	0.0	0.20609915966648532	0.0
Bruce Springsteen	0.0	0.0	0.0	0.0	0.7960272620964498	0.0	0.2039727379035503	0.0	0.0
Elvis Presley	0.0	0.0	0.0	0.0	0.0	0.0	0.4292247079977606	0.5707752920022394	0.0
Jimi Hendrix	0.24182961960319804	0.0	0.08607494935029084	0.0	0.0	0.0	0.17529327773124556	0.4968021533152656	0.0
Marvin Gaye	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
Michael Jackson	0.0	0.0	0.05685208539455243	0.023293562765823565	0.0	0.0	0.7010974145605853	0.218756937279038.7	0.0
Miles Davis	0.0974190283664351	0.0	0.8148523813361988	0.0284138832735435.7	0.0	0.0	0.03707775489670158	0.022236952127121058	0.0
Nirvana	0.0	0.5625562016665667	0.0	0.09375936694442778	0.29782387147053535	0.0	0.0	0.045860559918470115	0.0
The Beach Boys	0.0	0.0	0.0	0.0	0.298477267085111	0.0	0.5360625087698818	0.1654602241450072.2	0.0
The Beatles	0.353076246313812	0.0	0.051746967306211734	0.3574031610970695	0.0	0.0	0.0623642625042417	0.1754093627786649.2	0.0
The Clash	0.0	0.0	0.0	0.009833351463003806	0.29152994925611286	0.6909410329185326	0.0	0.00769566636235080.5	0.0
The Rolling Stones	0.0	0.0	0.0	0.0	0.0	0.0	0.40404877387760363	0.5959512261223964	0.0
The Velvet Underground	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
Van Morrison	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

Summary

PROS

Integration of cultural metadata on MIR research

Flexibility in filtering, limiting and presenting results

Extensibility of the API to new databases

Summary

CONS

Complex filtering: time & query consuming

Query sensibility

Distortion results

Future

Databases

Natural Language

P2P / Streaming Services

Applications

Databases

Automated access to several
distributed databases on the web

Use tags and user information
from applications and librarys

Development of site weighting and
filtering according to categories and queries

Natural Language

to query and retrieve information
about music and artists

to extract tags

to convert raw information to metadata

focus in common musical language

P2P and stream

use professional and user knowledge :
extracting playlists from online radios and mixtapes

study music transfer on P2P nets

Applications

TOP100 Web
(downloads and stream)

track genre evolution over time

Bibliography

http://www.music.mcgill.ca/~gabriel/courses/mumt621/presentations/4_jWM/annotated_bibliography.html

McKay, C., and I. Fuchinaga. 2007. jWebMiner: a web-based feature extractor. In *Proceedings of the 8th International Conference on Music Information Retrieval*. 113-4.

McKay, C., and I. Fuchinaga. 2008. Combining features extracted from audio, symbolic and cultural sources. In *Proceedings of the 9th International Conference on Music Information Retrieval*. 597-602.

Whitman, B., and P. Smaragdis. 2002. Combining musical and cultural features for intelligent style detection. In *Proceedings of the 3rd International Conference on Music Information Retrieval.* 47-52.

Whitman, B., and S. Lawrence. 2002. Inferring descriptions and similarity for music from community metadata. In *Proceedings of the 2002 International Computer Music Conference*. 591-8.

Zadel, M., and I. Fuchinaga. 2004. Web services for music information retrieval. In *Proceedings of 5th International Conference on Music Information Retrieval.* 478-83.

Thanks !