

Képi és szemantikus keresőalgoritmusok kutatását támogató közösségi platform

Gergő Márton, Havasi László, Mátételki Péter, Micsik András, Kovács László, Szirányi Tamás
MTA SZTAKI, 1111 Budapest, Lágymányosi utca 11.

marton.gergo@dsd.sztaki.hu, {[havasi](mailto:havasi@dsd.sztaki.hu), [matetelki](mailto:matetelki@dsd.sztaki.hu), [micsik](mailto:micsik@dsd.sztaki.hu), [kovacs](mailto:kovacs@dsd.sztaki.hu), [sziranyi](mailto:sziranyi@dsd.sztaki.hu)}@sztaki.hu

Kivonat: Az új keresési megoldások kifejlesztése sok kísérletezést igényel, és az ezzel kapcsolatos tevékenységet kívánjuk egy e-science közösségi platformmal támogatni. A bemutatott megoldás keretet ad többféle keresési algoritmus egyidejű kipróbálására, kombinálására, valamint a tesztadatok megosztására. A platform részeként a SZTAKI-ban fejlesztett újfajta képkeresőt és szemantikus keresőt is bemutatjuk.

Bevezetés

Mind a képi, mind a szemantikus keresőalgoritmusok fejlesztéséhez megfelelő méretű és minőségű adatbázisra, az adathalmazokon számos tesztet futtatására, s az egyes megoldások eredményeinek összehasonlíthatóságára van szükség. Új e-science platformunk célja a képi és a szemantikus keresőalgoritmusok iteratív fejlesztési folyamatának támogatása, a kutatói csoportok kollaboratív működésének segítése. A CrossMedia rendszerben lehetőség nyílik a keresések fejlesztésére felhasznált adathalmazok kollaboratív építésére, bővítésére, azok különböző szinteken történő megosztására. Továbbá a rendszer lehetőséget biztosít a kutatók által készített képi feldolgozó-algoritmusok saját, elosztott rendszerünkben történő futtatására, ezáltal az ún. indexfák építését webszolgáltatásként kínálja. A portál az így előállított, kereshető adatstruktúra teszteléséhez rugalmas keresőfelületet biztosít, továbbá az előállított indexfák a külvilággal is megoszthatóak, így kutatói csoportok egymás eljárásait is tesztelhetik, eredményeiket véleményezhetik.

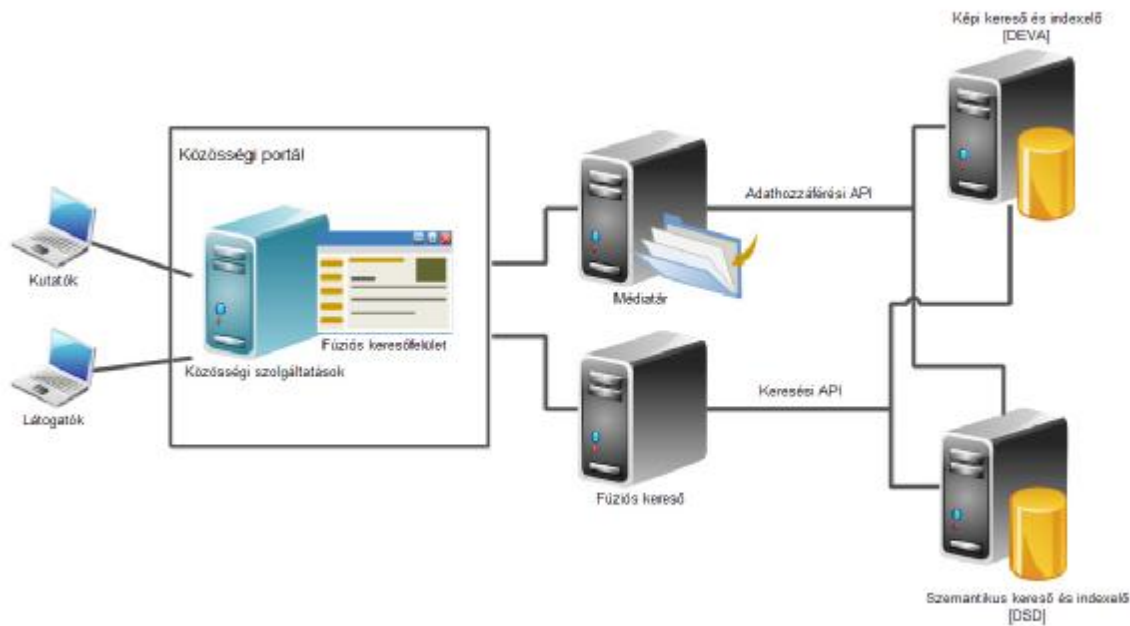
A rendszer fejlesztésével párhuzamosan saját, mind képi, mind szemantikus keresési eljárások fejlesztésére is sor került. Ezen eljárások eredményeképp az adathalmazokon akár multimodális keresési feltételek is értelmezhetővé válnak. A portál részeként ezen fúziós keresési eljárások is bemutatásra kerülnek. A platform fejlesztése a SZTAKI-n belül az Elosztott Rendszerek Osztály illetve az Elosztott Események Elemzése Kutatólaboratórium együttműködésében valósult meg.

A keretrendszer

A keretrendszernek számos, eltérő feladatot kell összehangoltan megoldania, amely feladatok megoldása bizonyos esetekben különböző kutatólaborok specifikus tudását is igénylik. Emiatt a kifejlesztett rendszer egy elosztott, számos feldolgozóegységből álló architektúrára épül. A rendszer részei:

- A képi kereső motor és a képi indexelő, amely a Médiatárba megfelelő algoritmusokkal a megfelelő indexekbe tölti a képeket, illetve azokon keresést végez

- A szemantikus kereső és indexelő, amely a Médiatárba töltött képek szabad szöveges annotációi alapján szemantikus adatbázist épít, és keresést valósít meg
- A fúziós kereső, amely az előző két kereső eredményeit összegzi
- A médiatár, amely a kereshető média-elemeket tárolja
- A közösségi portál, amely közösségek létrehozását, és a kollaboratív munkát támogatja
- A portállal integrált fúziós kereső felület

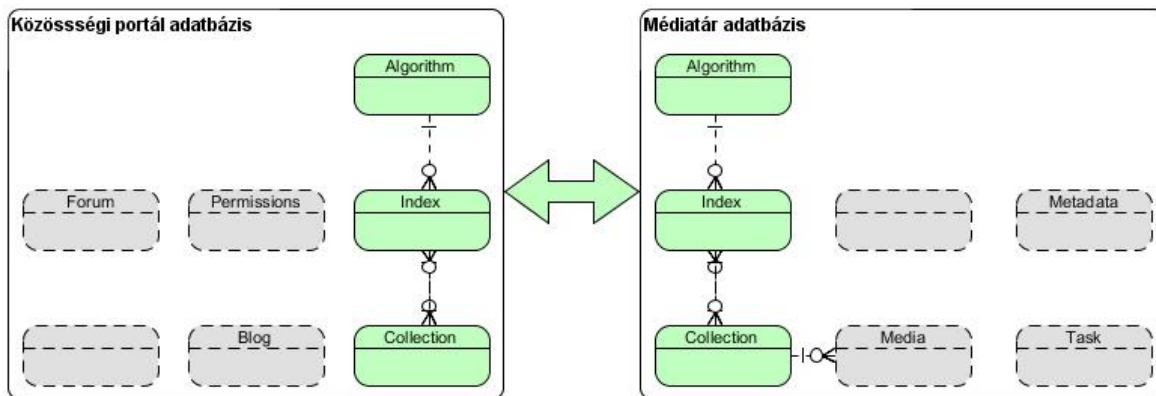


1. Kép: A CrossMedia keretrendszer elosztott architektúrája

Adathalmazok menedzselése

A fejlesztés során az egyik legnagyobb kihívást a nagy mennyiségű adat kezelése okozta. A keresési eljárások hatékony teszteléséhez kifejezetten nagyméretű adathalmazokra van szükség - a saját eljárásaink tesztelésére használt adathalmaz körülbelül 5 millió entitást tartalmaz. A platform célul tűzte ki ezen adathalmazok kollaboratív építésének, karbantartásának támogatását, azonban számos jelenlegi rendszer evaluációja után világossá vált, hogy ekkora adathalmazok megosztott menedzselési és jogosultságkezelési kérdései nem triviálisan megoldhatóak. Ez saját architektúra kialakítását, s fejlesztését tette szükségessé.

A menedzselési problémák megoldásához különválasztottuk a rendszer közösségi funkcióit az adattárolás feladataitól. Az adatok tárolását egy alacsony tudású, ám gyors Médiatár végzi, míg a közösségi jellegű feladatok a portálrendszer hatáskörében maradnak. Ennek a megoldásnak egyik előnye az, hogy az adathalmazok viszonylag függetlenek az azokat menedzselő front-end rendszertől, így meghagyva a lehetőségét a portál rugalmas átalakítására. Másik előnye, hogy a megfelelő menedzselési feladatok működéséhez csupán néhány Médiatárbeli elem másolatára és szinkronban tartására van szükség portál oldalon, így a rendszer nagy adatmennyiség mellett is skálázható marad. A Médiatár az adatok tárolásán túlmenően megteremti a kapcsolatot a közösségi oldalon létrehozott entitások és a médiafeldolgozó-egységek között. A Médiatár és a különböző feldolgozóegységek REST API-n keresztül kommunikálnak.



2. Kép: A Médiatár és a Közösségi portál közös entitásai

Multimodális instant keresési protokoll

A kidolgozott Keresési API egységesen kezeli a különböző modalitású keresési feltételeket, s támogatja az összetett, akár eltérő modalitású keresési paraméterekből összeállított komplex kérések létrehozását is. Ennek célja egyrészt a későbbi könnyű bővíthetőség, másrészt újszerű, multimodális keresések megvalósítása. Az így kialakított komplex kérések kiértékelését egy speciális, Fúziós kereső végzi, amely a kéréseket a megfelelő keresőkhöz továbbítja, majd a bejövő találati listák aggregálása során egyetlen megjelenítendő listát állít elő. A Fúziós kereső kialakítása révén részeredményeket is képes szolgáltatni, így segítségével instant keresőfelület építhető, amelyen részleges találati lista a kérés elküldése után röviddel megjeleníthető, s a keresési paraméterek változtatásának hatása rövid késleltetéssel érzékelhető. A találati listán szereplő elemekről a felhasználónak lehetősége van visszacsatolást létrehozni azok megfelelő vagy helytelen találatként értékelése révén. A visszacsatolás segítségével a találati lista fuzionálásakor használt „súlyfüggvény” finomhangolása megy végbe, amely segítségével hosszútávon a keresési eredmények javulását várjuk.

A közösségi portál

A portál elsődleges célja a kutatók kollaborációjának segítése területspecifikus igényeik figyelembevételével. A portálra regisztrált felhasználók kutatói csoportokat alkotnak. A csoporton belüli kommunikációt ismert, bevált eszközökkel támogatjuk, úgy mint beszélgetés-folyamokat kezelő fórum, gazdag hozzászólási lehetőségek, aktivitások követése.

A csoporthoz tartozó térben válik lehetségessé médiakollekciók létrehozása, azok kollaboratív építése. Lehetőség nyílik a kollekciónban lévő képek szemantikus annotálására, így a képek között szemantikus keresés valósítható meg. Továbbá lehetőség nyílik saját algoritmusok feltöltésére, megosztására. Az algoritmusok segítségével indexek definiálhatók, amelyekhez kollekciónk rendelhetők, így hozva létre a képi kereséshez használt indexstruktúrákat. A fejlesztési folyamat támogatása az ezen összerendelés során létrejövő feladatok automatizált elvégzése révén valósul meg.

Az indexek kollekciónkkal történő összerendelésre a háttérben számos feladat elvégzése szükséges. Az esemény során a médiatárba kerülnek az elvégzendő feladatokat leíró adatok, s a back-end elvégzi a szükséges állományok megfelelő algoritmusokkal történő leírókinyerését, s

indexelési feladatait. A gyakran időigényes folyamat állapotáról állandó visszajelzést kaphat a felhasználó. A kollekciók utólagos bővítése, módosítása, szintén indexfrissítési feladatokat generál, amelyeket a rendszer automatikusan elvégez, így a kutató naprakész indexeken tesztelheti eljárásait. A kollekciók különböző szinteken, akár csoportok között is megoszthatók, így teremtve meg a lehetőséget, hogy különböző csoportok saját eljárásaikat egyazon adathalmazon összehasonlítsák.

A kifejlesztett képleíró és szemantikus keresési indexek tesztelésére egy jól áttekinthető, könnyen és intuitívan kezelhető keresőfelületet hoztunk létre. A felület biztosítja az elérhető média-adatbázison az indexekre vonatkozó összetett keresési feltételek megfogalmazását és az összefésült találati eredmények megjelenítését. A keresőfelületet böngésző- és platformfüggetlen web-alkalmazásként implementáltuk a Sencha platform segítségével, működése teljes mértékben Ajax technológiára épül. Az indexek és a kollekciók külvilág felé is megoszthatók, melynek révén az eljárások külső látogatók számára is elérhetővé válnak. Így a portál demonstrációs felületként is szolgálhat a kutatói csoportok számára, illetve a publikációs tevékenységet is támogatja.

Saját fejlesztésű keresési eljárások

Demonstrációs célokra a keretrendszerbe töltöttünk körülbelül 5 millió képet a CoPhIR (Content-based Photo Image Retrieval) adatbázisból [1]. Az adatbázis flickr-ről letöltött képek metaadatait, továbbá a képekről kinyert MPEG-7 leírókat tartalmazza.

Szemantikus adatbázis

A szemantikus adatbázis építése során a DBpedia által biztosított ontológiát használtuk [2]. A CoPhIR adatbázisban található metaadatok elemzése során kiderült, hogy a képekhez tartozó szöveges információ igen zajos és szegényes, a kép tartalmáról gyakorlatilag csak a hozzájuk tartozó címkék (tag-ek) hordoznak érdemi információt. A szemantikus feldolgozás során ezeket a címkéket entity extraction módszerek alkalmazásával DBpedia node-okkal társítottuk. Az előállított szemantikus társítások alapján célzott szemantikus keresések megvalósításával kísérletezünk, mint például napszakok szerinti keresés. Általánosan a kategóriák szerinti keresést is megoldottuk oly módon, hogy a szemantikus relációk tranzitivitását kihasználva az alkategóriák is megtalálhatók. Például ha a Trains kategóriában keresünk, akkor olyan képeket is megtalálunk, amelyhez az "eurostar" vagy "shinkansen" címke van hozzárendelve.

Index struktúrák a képi kereséshez

Képtartalom alapú kereséshez a CoPhIR adathalmazban rendelkezésre álló MPEG-7 globális képleírókat alkalmaztuk. Ezek a leírók a teljes képet jellemzik egyetlen nagy dimenziós vektorral, melyek alapvetően a színre, vagy a mintázatra adnak jellemzést. Tapasztalataink alapján a color-structure, edge histogram és a homogenous texture leírók [6] a leginkább alkalmasak a képtartalom szerinti illeszkedés keresésére.

A nagy mennyiségű és nagy dimenziószámú adatok kereshetővé tétele speciális adatstruktúrát igényel, főbb szempontok a következők:

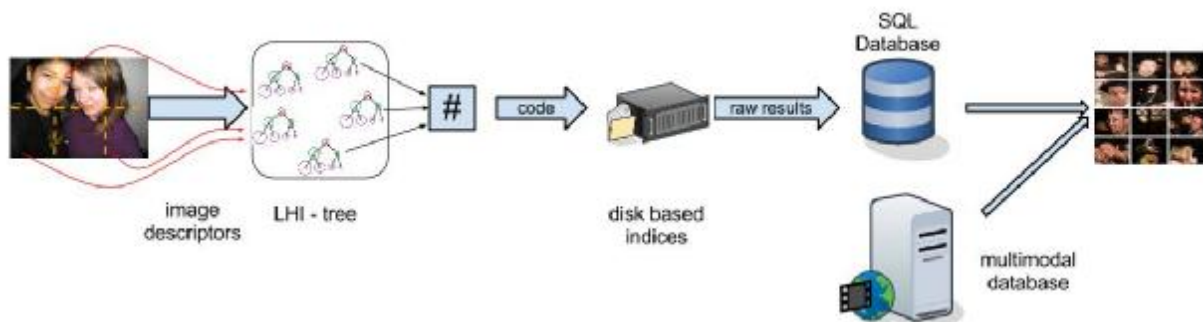
- nem vektoriális adat reprezentáció támogatása: számos eljárás épít a leírók fix dimenziójú, vektoriális adatrepresentációjára és az ezen értelmezhető, elsősorban

statisztikai algoritmusokra. Ezzel szemben számos state-of-the-art tartalom (elsősorban képrészlet) leíró változó méretű és jelentésű adatsomagokból épül fel.

- nem metrikus távolság definíció támogatása: az index struktúrák jelentős része kizárólag metrikus terek esetén alkalmazható. Ellenkező esetben futási és indexelési hibákkal leáll az index építési folyamat. Célunk, hogy a metaheurisztikus, tapasztalati úton kialakított távolság definíciók is alkalmasak legyenek az index felépítésére [4].
- lemez alapú adattárolás: nagy mennyiségű adatok esetén hamar elérhetjük a memória kapacitásának határait. Bővíthető adatstruktúra esetén elengedhetetlen, hogy az indexek is a lemezen kerüljenek tárolásra. Ennek a követelménynek a legtöbb memória használatra optimalizált keresési fa nem felel meg, mivel nagy számú lemezműveletet igényelne ami használhatatlanul lassú keresést eredményez. Optimális esetben egyetlen lemez olvasás a megengedett [3], melynek alapját a memóriában elhelyezhető segédindexek adják.

A követelményeknek megfelelő struktúra kiindulási pontja a teret reprezentáló, lehetőleg ortogonális bázisrendszer felállítása. Esetünkben ez 5 bázis kiválasztását jelenti, melyek az alapját adják a referencia távolság alapú hash-kód számolásnak. Új elemek beszúrása a bázisoktól vett távolság kvantálásával nyert hash-kódok alapján lehetséges, melyek egy-egy metszetet határoznak meg a nagy dimenziós térben. Ezek a metszetek a háttértáron állományoknak felelnek meg melyek tartalmát a metszetbe eső minták azonosítói adják.




Keresés során a bemenetet jelentő minta leíró alapján meghatározásra kerül a tér egy metszete melynek bővítését az eltérő skálák szerinti struktúrák párhuzamos használata tesz elérhetővé. A metszet tartalmának közelítő rendezésével nyert lista adja a találati listát ami az N legközelebbi elemet fogja tartalmazni.






3. kép: Keresési folyamat szemléltetése.

	3sphs/300k 64dims	5sphs/600k 64dims	5sphs/829k 80dims	5sphs/632k 128dims
Build time (sec)	456	12194	17619	14901
RAM storage (MB)	0.912	1.248	1.432	0.945
HDD storage (MB)	80200	161000	215000	329210
Search time (msec)	67	45	110	160
RAM usage during search (MB)	1.2	0.97	1.2	2.2

4. Kép: Index építési és keresési eredmények eltérő dimenziószámú leírók esetén

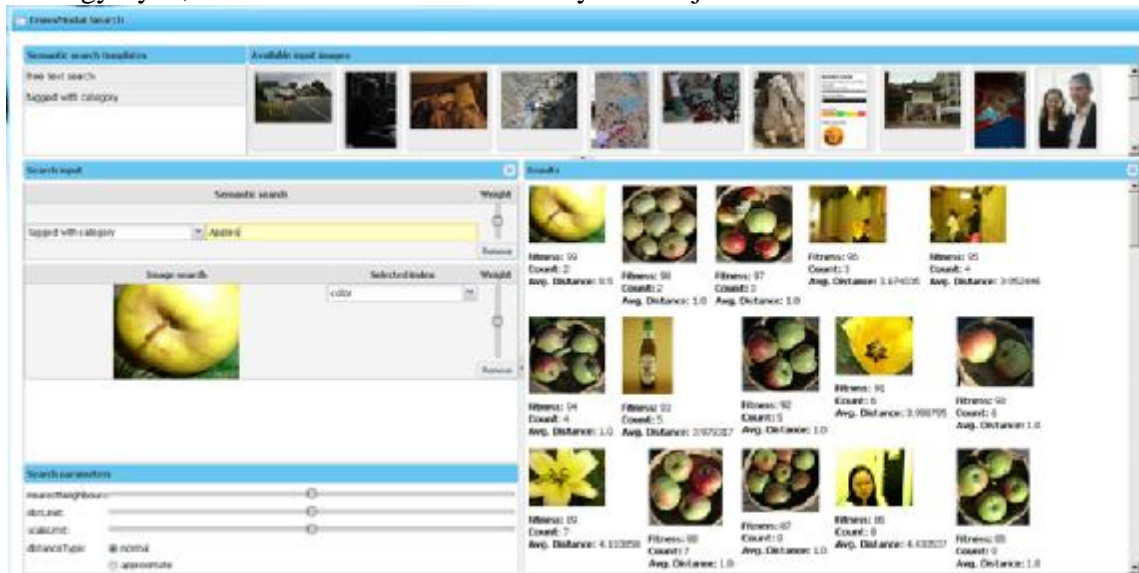
Color structure descriptor	
	Descriptor length is 64 dims.
FLANN	LHI-tree
	

5. Kép: Találati lista összehasonlítása globális képleíró esetén, referencia listát a nagy pontosságú memória alapú FLANN index adja.

SIFT features without dimension reduction, pattern difference for LHI-Tree and Euclidean metric for FLANN	
	Specific object category without background content. Total 211 descriptors in the image.
FLANN	LHI-tree
	

6. Kép: Találati lista összehasonlítása régió alapú képleíró esetén, referencia listát a nagy pontosságú memória alapú FLANN index adja [5].

A kialakított eljárások összehangolása révén újszerű, multimodális keresések megvalósításával próbálkozunk. A kialakított felületen a keresési feltétel bemeneteként akár több képet, vagy több szemantikus kifejezést is megadhatunk, s megvizsgálhatjuk a paraméterek hatását a találati listára. Egy ilyen, multimodális keresés eredményét mutatja az alábbi ábra:



7. Kép: Az eltérő modalitású eljárások erősítik egymást, bizonyos képek szemantikus annotáció alapján, míg mások képi hasonlóság alapján kerültek a találati listába

Összefoglalás

A CrossMedia projekt során kutatásokat végeztünk a képi- és szemantikus keresések területein, mely során új indexstruktúrákat használó képi keresési eljárásokat, valamint kombinált szemantikus keresési eljárásokat alakítottunk ki, melyekhez a kifejlesztett e-science portál segítségével más kutatók is hozzáférhetnek és a további kutatásokban részt vehetnek. A képi keresés mintahalmazaként a kb. ötmillió képet tartalmazó adatbázist használtuk, mely során a nagy adatmennyiség által támasztott követelményeknek is meg kellett felelnünk. Az ismert memória-alapú képkeresőkhöz hasonló sebességű, de lemez-alapú indexelési algoritmusokat hoztunk létre, ezzel lehetővé téve a nagyméretű képadatbázisok használatát. A szemantikus és képi kereséseket kombináltuk: létrehoztunk egyrészt egy fúziós keresőt, mely képes a szemantikus és képi keresés során előállított eredmények összefésülésére, illetve létrehoztunk egy intuitívan használható multimodális keresőfelületet, melynek segítségével a felhasználók összetett keresési kifejezéseket képesek megfogalmazni. Ez a multimodális keresőfelület kiválóan használható az egyes algoritmusok tesztelésére, az algoritmusok összetett tesztelésére, valamint a keresési terek kombinálására is.

Az eredmények demonstrálására és publikálására olyan nyíltan használható e-science platformot építettünk, melynek felhasználásával más kutatói közösségek munkáit is képesek vagyunk támogatni. Az létrehozott portál segítségével a kutatók kollaborációs csoportokat alakíthatnak, melyben a szokásos portálfunkcionalitáson túl olyan szolgáltatásokat találnak, amik a mindennapos kutatási tevékenységeiket segítik elő: médiakollekciók létrehozása, keresési algoritmusok futtatása, indexstruktúrák létrehozása. Ezekon felül a portál lehetőséget biztosít a csoportok közötti együttműködésre, valamint az eredmények megosztására a külvilág felé.

Hivatkozások

- [1] Paolo Bolettieri, Andrea Esuli, Fabrizio Falchi, Claudio Lucchese, Raffaele Perego, Tommaso Piccioli, Fausto Rabitti. CoPhIR: a Test Collection for Content-Based Image Retrieval. CoRR abs/0905.4627: (2009)
- [2] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, Sebastian Hellmann: DBpedia – A Crystallization Point for the Web of Data. Journal of Web Semantics: Science, Services and Agents on the World Wide Web, Issue 7, Pages 154–165, 2009.
- [3] Lejsek, H.; Asmundsson, F.H.; Jonsson, B.T.; Amsaleg, L. : NV-Tree: An Efficient Disk-Based Index for Approximate Search in Very Large High-Dimensional Collections, Pattern Analysis and Machine Intelligence, IEEE Transactions on 31, (5), May 2009, 869 - 883
- [4] Horst Eidenberger, Distance measures for MPEG-7-based retrieval, in MIR 2003
- [5] C. Silpa-Anan, R. Hartley, Optimised KD-trees for fast image descriptor matching, in CVPR 2008
- [6] Mpeg-7, ISO/IEC JTC1/SC29/WG11N6828, October 2004, <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>