

Önálló laboratórium beszámoló

Dolgozat címe:

.....

Konzulens(ek) neve:

(Külső cég neve:

címe:

A Hallgató a kitűzött feladatot megfelelő színvonalon és a kiírásnak megfelelően

teljesítette

nem teljesítette

Konzulens aláírása

Hallgató neve:

Képzés:

Leadás dátuma:

Tartalomjegyzék

1. A feladat rövid ismertetése, célkitűzés.....	3
2. Bevezetés.....	4
2.1 Előzmények.....	4
2.2 Irodalmi áttekintés.....	5
2.2.1 PCA	5
2.2.2 WEKA.....	6
2.2.3 PostgreSQL	7
3. Eszközök és módszerek.....	10
3.1 Megvalósítás.....	11
4. Eredmények.....	15
5. Következtetések.....	15
6. Köszönetnyilvánítás	16
7. Irodalomjegyzék.....	16
8. Összefoglalás.....	17

Ábrajegyzék

1. ábra PostgreSQL pillanatkép.....	9
2. ábra Adatbázis felépítése.....	10
3. ábra GSM tábla általam használt értékei.....	11
4. ábra Pivot előtti adatok.....	12
5. ábra Pivot utáni adatok.....	12
6. ábra WEKA konfigurálás	13
7. ábra Vágókép a WEKA programból	13
8. ábra ARFF-SQL konvertálás.....	14
9. ábra Színezett adatok táblája	14

1. A feladat rövid ismertetése, célkitűzés

Önálló laboratóriumi munkám célkitűzése az volt, hogy önálló vizsgálatokkal, mérésekkel alátámasztva keressek egy optimális adatbázis-tömörítési eljárást, amellyel egyszerűbben tudunk dolgozni az adatbázisokon és jobban tudjuk reprezentálni a vizsgálni kívánt számszerűsített adatokat. A tömörítés lényege az, hogy az adatbázisban a kívánt adatok a lehető legtömörebbek, leg-helytakarékosabbak legyenek és a lehető legkönnyebben lehessen velük dolgozni.

Ezzel a módszerrel időt és helyet takarítunk meg, ezzel könnyebbé téve az adatok kezelését, az adatbázis jobb megismerhetőségét, összességében az adatbázison dolgozó szakember munkáját.

A félév során a Pázmány Péter Katolikus Egyetem, Információs Technológiai Karának GSM adatbázisán végeztem klasszifikációs eljárást. Fő feladatomban az volt, hogy teszteljem a PCA tömörítési eljárást az egyetem GSM adatbázis szerverén.

Önálló laboratórium 1 című tárgyam keretein belül, feladatomban volt a metodikai betanulás, ezen belül is az adatgyűjtés az adott témában, adatértelmezés, ezen ismeretek megfelelő felhasználása a gyakorlati munkám során, illetve a precíz dokumentálás. Irodalomkutatásom során, internetes források alapján megismertem a helyes hivatkozás, utalás, idézés és a lábjegyzet készítés formai követelményeit. Irodalomkutatást végeztem az adott témakörben, használva internetes forrásokat és elektronikus jegyzeteket. Információt kerestem a létező adatbázis tömörítési eljárásokról.

Megismertem a gyakorlati munkám során alkalmazott WEKA széles körben elterjedt, világszínvonalú adatbányászati programot és a PostgreSQL adatbázis nyelvet egyaránt. Végül behatóbban megismerkedtem a PCA tömörítési eljárással magával és ezzel végeztem vizsgálatokat. Eredményeimet a dolgozat végén összegzem.

A félév célja, a PCA eljárás, a WEKA adatbányászati program és a PostgreSQL nyelv megismerése, jövőben elkészítendő diplomamunkám tudományos igényességű előkészítése volt, hogy felkészítsen a következő önálló feladatomra, ami maga a diploma dolgozat elkészítése. Ehhez ez a félév megadott minden ismeretet és előképzettséget. Önálló laboratóriumi munkámat koordinálta és felügyelte Tihanyi Attila.

2. Bevezetés

Jelen dolgozatomban egy mobil turisztikai alkalmazás fejlesztésével foglalkoztam, amely az egyetemen implementált GSM alapú helymeghatározásra épül. A helymeghatározási feladatokat a munkacsoportban egymás között szétosztottuk. Az egész rendszer egy GPS alapon felvett referencia hálózat és egy GSM vételi jelszint térkép összevetésére alapul. Az adatgyűjtés során különböző eszközökből és különböző módokon keletkezett adatok kerültek egy „mobilhely” nevű adatbázisba. Az adatbázis GSM adatainak felhasználásával, kell módszert találni a fizikailag közel eső pontok keresésére. Mivel az adatbázis már jelenlegi formájában is sok rekordot tartalmaz, ha megtalálunk egy jól használható és paraméterezhető „távolság” fogalmat leíró rendszert, annak segítségével pozíciót meghatározni is lehet de alkalmas az adatbázis tömörítésére is. Feladata megközelítése során kihasználjuk a GSM terjedésről tanultakat, az analízis ismeretek alapján megpróbálunk veszteséges tömöríteni valamilyen ismert módszerrel, és a megkívánt eredmény felől közelítve előre meghatározott mérési ponthalmazok között teszünk különbséget. A saját módszereink és mások eredményeinek összevetéséből határozzuk meg a tovább követendő irányt.

2.1 Előzmények

A mai világban egyre gyakrabban használunk helymeghatározást különféle helyekhez jutáshoz, kereséséhez, informálásához. Fejlett technológiáknak köszönhetően egyre elterjedtebbek az olyan rendszerek, eszközök, amelyek a pozíció meghatározására épülnek. Ez általában GPS alapon történik, melynek lényege hogy több műhold segítségével lekérdezzük az aktuális helyzetünket.

Ehhez szükségünk van egy GPS alapú készülékre, illetve legalábbis egy vevőre. Ennek segítségével kérdezzük le az aktuális pozíciónkat. Előnye, hogy nagyon pontos, viszont szükséges a „szabad” rálátás a műholdakra. Ez sajnos nem mindig biztosítható. Továbbá nem biztos, hogy hétköznapi tárgyaink közt van olyan készülék, ami képes GPS funkciót ellátni. Ezen okok miatt fejlődött ki a GSM alapú helymeghatározás, mely hasonló elven működik. Itt a bázisállomás informál minket és a telefonkészülék a vevő. Ennél az eljárásnál már nem gond, ha fedett helyen szeretnénk helymeghatározást végezni.

Feladatomban kaptam, hogy vegyem számba a lehetőségeket a GSM adatbázis adatainak tömörítésére azzal a céllal, hogy megvizsgáljam különböző pontok közös információ tartalmát. Erre azért van szükség, hogy megoldást találjunk az adattovábbítás során fellépő felesleges információáramlásra, kiküszöböljük a szükségtelen információk küldését és fogadását.

2.2 Irodalmi áttekintés

A legelső lépés a szakirodalom áttekintésére a keresőszavak összegzése volt. Más szóval keresésoptimalizálás megfelelő kulcsszavakkal, szó-összetételekkel. A keresőszövegben speciális kulcsszavak használatával adhatunk meg összetettebb lekérdezéseket.

A stratégiám az volt, hogy táblázatba összegyűjtöm a keresőszavaimat annak megfelelően, hogy éppen melyik részt szeretném feldolgozni, keresőszó-csoportokat adtam meg a fejezetcímek alatt.

A szakirodalom feldolgozásakor az alábbi szakirodalmi hivatkozási fajtákat vettem számba:

1. internetes forrás
2. internetes folyóiratok
3. internetes adatbázisok

2.2.1 PCA

Kulcsszavak: PCA, principal component analysis, komponens analízis.

Munkám során a főkomponens analízissel végeztem vizsgálatot, amely olyan elemzési módszer, melynek célja: a megfigyelési egységek csoportjainak elemzése, háttérváltozók felderítése és így az eredeti megfigyelési változók közötti lényeges kölcsönhatások megállapítása.

Főkomponens analízis (PCA) egy matematikai eljárás, ahol a főkomponensek száma kisebb vagy egyenlő, mint az eredeti változók száma.

Dimenzió redukciós eljárás, amelynek során olyan transzformációt keresünk egy adatmátrixhoz, amelynek alkalmazása az alábbi tulajdonságokkal rendelkező transzformált adatmátrixot eredményezi:

Minden attribútum esetén 0 a kovariancia.

Az attribútumok annak sorrendjében rendezettek, hogy milyen mértékben járulnak hozzá a szóráshoz: az első attribútum járul hozzá legnagyobb mértékben a szóráshoz, az utolsó a legkevésbé.¹

PCA: dimenziócsökkentés tér transzformációval. Mivel a sajátvektorok variancia (amit a sajátérték ad meg) szerint vannak csökkenő sorrendben, meg lehet adni azt, hogy a szórás hány százalékát tartsuk meg transzformáció után. A kis szórású irányok (amelyekhez kis sajátérték tartozik) kevésbé informatívak, ezért azt elhagyhatjuk.²

2.2.2 WEKA

Kulcsszavak: WEKA, adatbányászati program, adatbányászat, adathalászat.

Weka, a valós adatbányászati problémák megoldására szolgáló algoritmusok gyűjteménye. Java alapú és fut szinte minden platformon. Képes kommunikálni egy adott adatbázissal, konfigurálás után. Ez munkám során bemutatásra kerül. Az algoritmusokat lehet alkalmazni, közvetlenül vagy meghívni Java kódból.³

A WEKA nevű adatbányászati szoftvert munkám elején töltöttem le az internetről, a lábjegyzetben megjelölt időben és forrásról.⁴

A WEKA nevű programcsomagot a University of Waikato munkatársai fejlesztették ki. (WEKA- Waikato Environment for Knowledge Analysis). Nevét egy Új-Zélandon őshonos, veszélyeztetett és röpképtelen madárról kapta, a Weka madárról.

¹ [Jeszenszky, 2011] <http://www.inf.unideb.hu/~jeszy/download/datamine/doc/pca.pdf> 2012.03.20. 12:20

² [Bánhalmi, 2010] www.inf.u-szeged.hu/~banhalmi/adatb/2_eloadas.ppt 2012. 03.12. 15:30

³ [WEKA, 2010] <http://sourceforge.net/projects/weka/> 2012.03.25. 09:15

⁴ [Software, 2012] <http://www.cs.waikato.ac.nz/ml/weka/> 2012.03.10. 19:45

2.2.3 PostgreSQL

Kulcsszavak: PostgreSQL, SQL, adatbázis nyelv.

A PostgreSQL, más néven *Postgres* egy relációs adatbázis-kezelő rendszer. Licencét tekintve szabad szoftver. Sok más szabad szoftverhez hasonlóan a fejlesztést önkéntesek végzik közösségi alapon.⁵

A PostgreSQL számos előnyt jelenthet a felhasználók számára.

Immunitás a túlterhelés ellen

A túlterhelés az a probléma, amit néhány kereskedelmi adatbázisszoftver-cég első számú licenelési problémája. A PostgreSQL-lel senki sem kérheti számon, mivel nincs szoftverlicenc-költség.

- jobban megtérülő üzleti modell nagy mennyiségű telepítés esetén
- nincs licencvizsgálat
- lehetőség, hogy próbafejlesztéseket hajtson végre vagy tesztrendszereket futtasson plusz költségek nélkül

Jobb terméktámogatás áll rendelkezésre, mint védett adatbázisszoftverek esetében: a támogatási ajánlatok mellett elérhető a PostgreSQL-szakértők és rajongók közössége, amely segítségét igénybe veheti vagy csatlakozhat hozzá.

Jelentős megtakarítások az adminisztrációs költségeken: az adatbázist úgy tervezték és építették meg, hogy sokkal kevesebb adminisztrációs és beállítási szükséglete legyen mint vezető kereskedelmi adatbázis-szoftvereknek, viszont így is biztosítja a teljesítményt, a lehetőségeket és a stabilitást.

⁵ [WIKI, 2010] <http://hu.wikipedia.org/wiki/PostgreSQL> 2012.04.05. 13:40

Megbízható és stabil

Más kereskedelmi adatbázissal ellentétben, a PostgreSQL-t választó cégek számára természetes, hogy az adatbázisuk soha nem áll le hibával akár évekig is működés közben. Egyetlenegyszer sem. Egyszerűen csak működik.

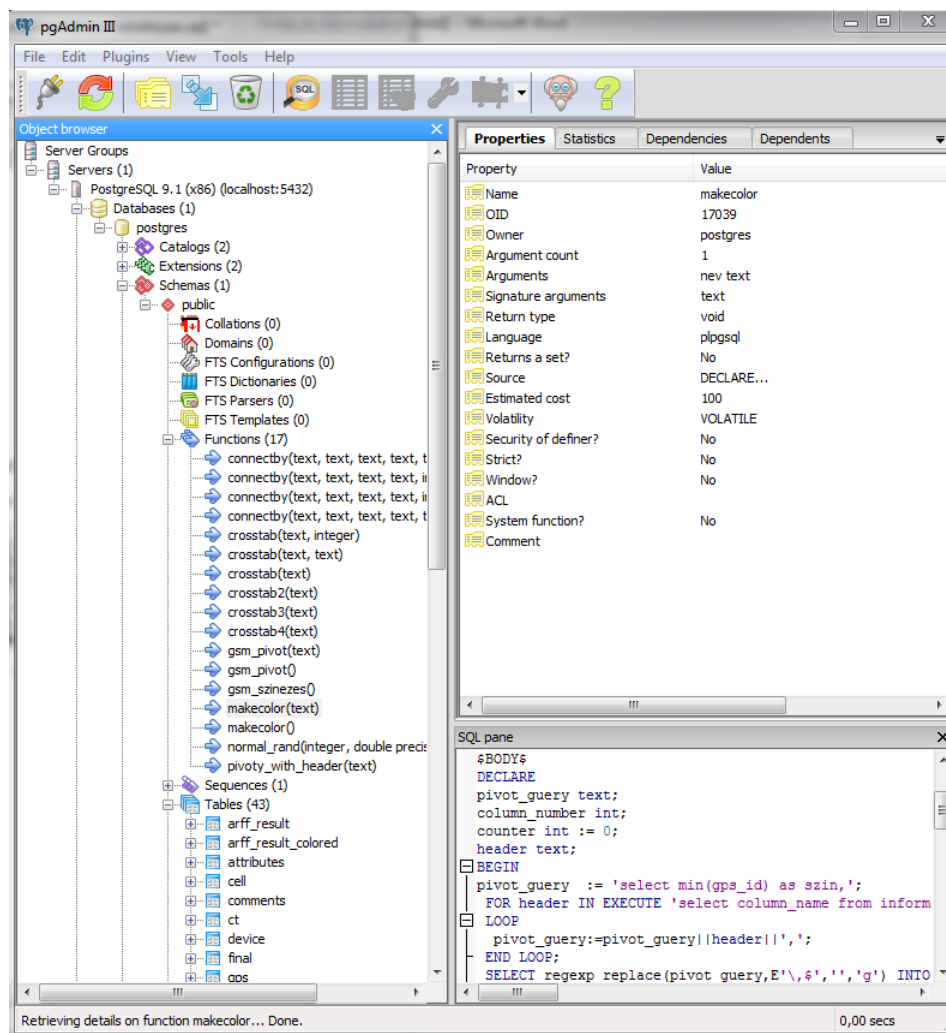
Kiterjeszthető

A forráskód szabadon hozzáférhető.

A PostgreSQL által kínált technikai lehetőségek:

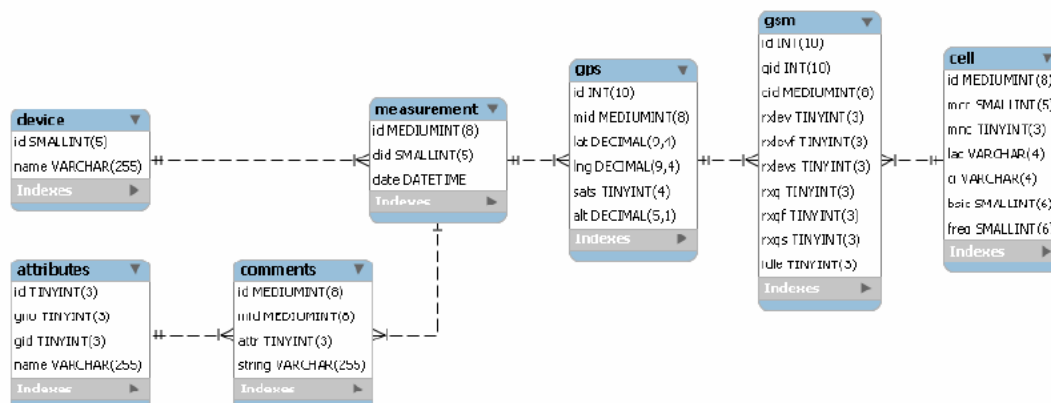
- tökéletesen megfelel az ACID szabványnak
- teljesíti az ANSI SQL szabvány kritériumait
- hivatkozási integritás (referential integrity)
- replikáció (kereskedelmi és szabad megoldások), a fő adatbázis mentése több gépre
- natív programozási felületek ODBC, JDBC, C, C++, PHP, PERL, TCL, ECPG, Python és Ruby programnyelvekhez
- szabályok (rule)
- nézetek (view)
- triggererek
- Unicode-támogatás
- szekvenciák
- öröklődés
- outer join
- al-szelekciók
- nyílt programozói felület
- tárolt eljárások
- natív SSL-támogatás
- procedurális nyelvek (tárolt eljárások)

- állandó készenlét (kereskedelmi megoldások)
- a sorsintű zárolásnál jobb mechanizmus
- betölthető SHA1, MD5, XML kiegészítők és egyéb funkcionalitás ⁶



1. ábra PostgreSQL pillanatkép

⁶ [HUPWiki, 2012] <http://wiki.hup.hu/index.php/PostgreSQL> 2012.04.10. 15:20



2. ábra Adatbázis felépítése⁷

Összesen 7 tábla tárolja a mérési adatokat. Ezek közül a félév során főleg 2 táblán dolgoztam.

- Gps, földrajzi helyzet meghatározás céljából.
- Gsm, mérések tulajdonságai.

3. Eszközök és módszerek

Munkám során, az egyetemen található GSM szerver adataival dolgoztam. Ezen végeztem különféle tömörítési eljárásokat tapasztalat szerzés céljából. Önálló munkámhoz saját laptopomat használtam. Eszközként megnevezném a munkámat nagyban segítő programokat, a WEKA adatbányászati programot és a Microsoft Office alkalmazásait. Legfontosabb lépésként említeném meg az irodalomkutatót, amivel az egész önálló munkámat kezdtem. Az általam még nem megfelelően ismert témakört ezzel a módszerrel ismerhettem meg a legjobban. Elolvastam számos cikket, írást, magyar és külföldi forrásokból, ezzel átfogóbb képet kaptam az engem érintő problémakörökről és az általam megoldandó feladatokról. Majd az irodalmi áttekintést követte egy részletes munkaterv kidolgozása, amihez próbáltam tartani magam. Majd következett maga a megvalósítás.

⁷[Majer, 2010] http://digitus.itk.ppke.hu/~tihanyia/HelyUltra/Majer_Julianna_Dipl.pdf 2012.04.14. 20:15

3.1 Megvalósítás

Legelső feladatom a félév során, az irodalomkutatás volt. Lényege, hogy megismerjem a lehetséges eszközöket és módszereket, amelyekkel az önálló laboratórium 1 tantárgy keretein belül meg tudom tervezni és végre tudom hajtani az általam vállalt féléves önálló munkámat.

Az irodalomkutatás után feltelepítettem egy PostgreSQL adatbázist, az adatok vizsgálata, használata céljából. Erre azért volt szükség, mert nem akartam a jól működő adatbázist a kezdeti próbálkozásaimmal veszélyeztetni. Továbbá jelentősen gyorsabb a lekérdezés helyi hálózatot használva. Megjegyezném, munkám teljes mértékben működőképes az egyetem adatbázisán is. A telepítés után, feltöltöttem az adatbázist az egyetemi szerveren található adatokkal. Munkám első jelentősebb állomása egy úgynevezett crosstab/pivot tábla megszerkesztése volt. Jelenleg az adatbázisban a GSM tábla tárolja a gsm mérés tulajdonságait.

	id integer	gid integer	cid integer	rxlev smallint
1	2	1	2	25

3. ábra GSM tábla általam használt értékei

- Az id mutatja a mérés sorszámát, ez különbözteti meg a többi méréstől.
- A gid értéke mutatja meg az adott mérés gps tulajdonságait. A mérés gidje egyenlő a GPS tábla id-val.
- A cid értéke mutatja a látott cellát. A mérés cidje egyenlő a CELL tábla id-val.
- Az rxlev értéke a mért vételi jelszint, a cid-ben meghatározott cellára.

Sajnos ezzel az a probléma, hogy egy adott mérésünk a GSM táblába felsorolva szerepel egymás alatt, megegyező gid-del. Így nem tudunk rendesen következtetni semmilyen algoritmussal sem, hisz az adataink szétdarabolva vannak eltárolva. Ebben nyújt segítséget a pivottolás. Lényege, hogy az azonos méréshez tartozó értékeket egy rekordba rendezi úgy hogy a látott cella száma alatt található meg a vételi jelszint. A pivot function futtatása előtt leredukáltam az oszlopszámot 1994-ről 1436-ra, mivel a PostgreSQL maximálisan 1600 oszlopot enged meg.

	gid integer	cid integer	rxlev integer
1	595	74	20
2	595	70	23
3	595	79	15

4. ábra Pivot előtti adatok

	gps_id integer	cid_8 integer	cid_9 integer	cid_70 integer	cid_74 integer	cid_79 integer
1	595			23	20	15

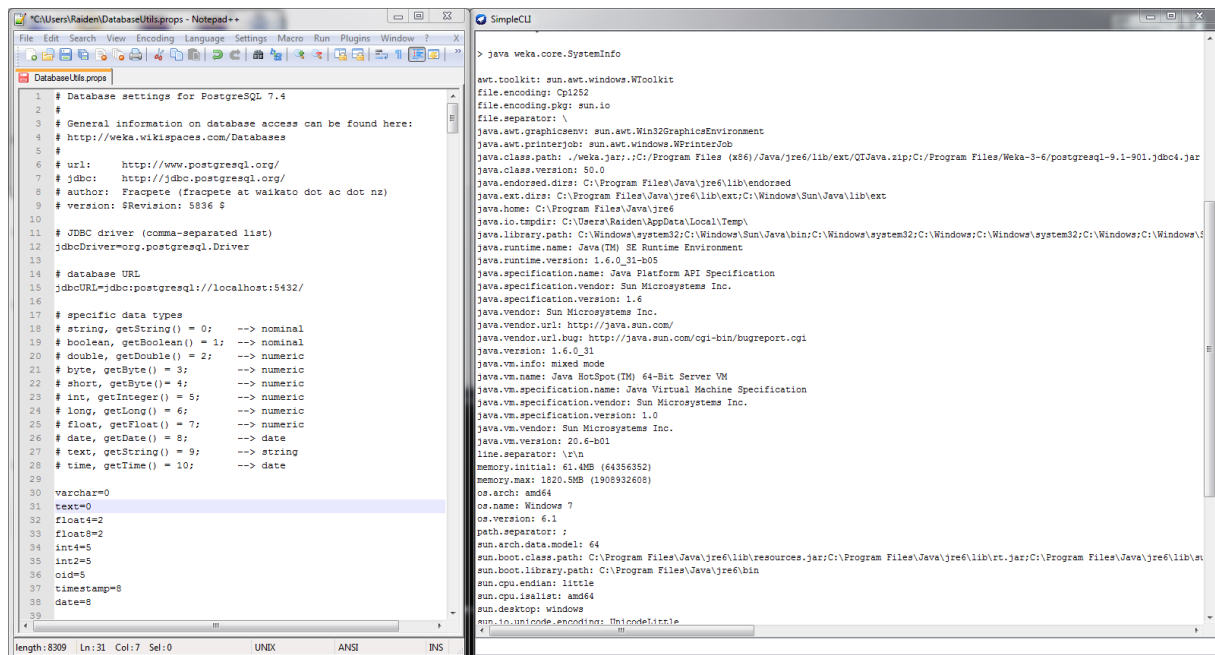
5. ábra Pivot utáni adatok

Ezt követően feltelepítettem a WEKA nevű adatbányászati programot. Majd közelebbről is megismertem, tanulmányoztam a program megfelelő alkalmazásait, lehetőségeit. Itt értem el munkám második jelentősebb állomásához. Be kellett konfigurálnom a WEKA-t, hogy képes legyen hozzáférni az adott szerverhez, esetemben a helyi adatbázishoz. Első lépésként letöltöttem a lábjegyzetben megjelölt forrásból az adatbázissal megegyező drivert.⁸ Ezt követően beállítottam a gépemen, hogy a Classpath tartalmazza a letöltött fájlt is. Ezt hívja meg a program induláskor. A konfigurálás után a program érzékeli a drivert. Viszont még szükséges a databaseutil.props fájl bemásolása a Users mappába, mivel ezt a WEKA home directory-ja. A databaseutils.props fájl a WEKA mappa(telepítés helye) weka.jar fájlában található meg, databaseuti1s.props.xxx-ként, ahol az 'xxx' az általunk használt adatbázis, esetemben postgresql. Mikor a fájl a helyére került, ki kell egészíteni ezzel a sorral:

jdbcURL=jdbc:postgresql://localhost:5432/ Ezzel definiáltuk a servert, portot.

Ezek után a WEKA képes kapcsolódni a szerveremhez, lekérdezni az adatbázisból. Sajnos ezzel még nem értem el a kívánt célt, mivel a WEKA alapértelmezésben nem ismeri az int4 típust, ezért ezt is definiálni kell a fent említett databaseutils.props fájl specific data types részében. A lépés után, már képes voltam kapcsolódni és lekérdezni az adatbázisomból. Utolsó lépésként, megnőveltem a virtuális memória nagyságát, a nagy adathalmaz miatt. Ezt a java -Xmx2048m parancs kiadásával értem el, a Simple CLI konzolban.

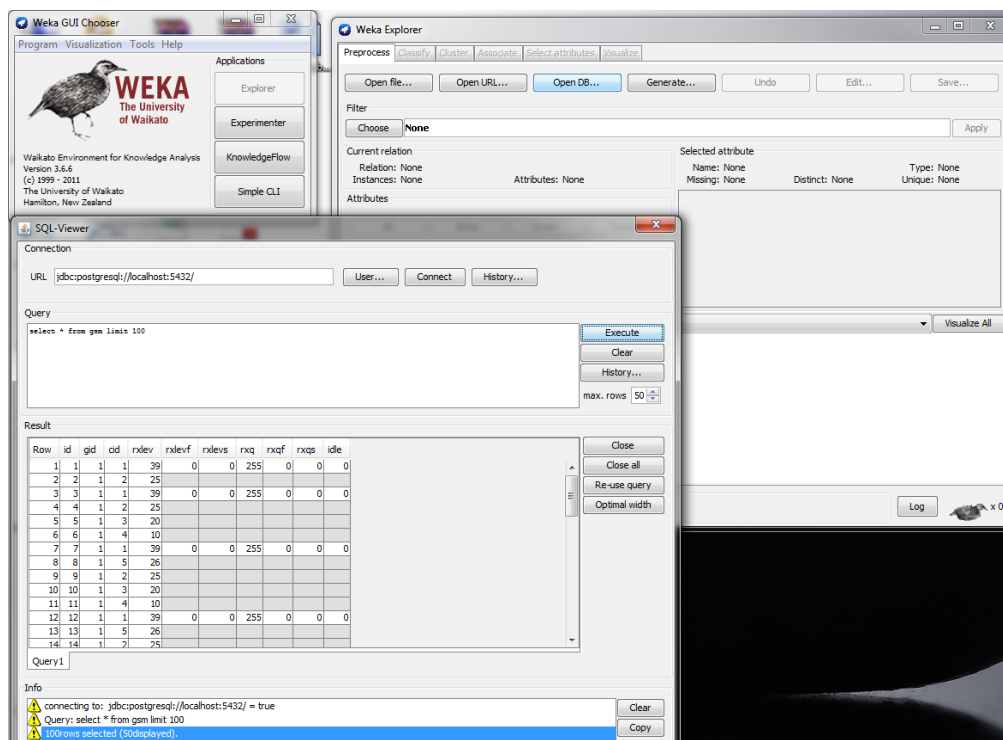
⁸ [WEKA, 2012] <http://weka.wikispaces.com/Databases#Setup-Driver> 2012.04.21. 22:45



6. ábra WEKA konfigurálás

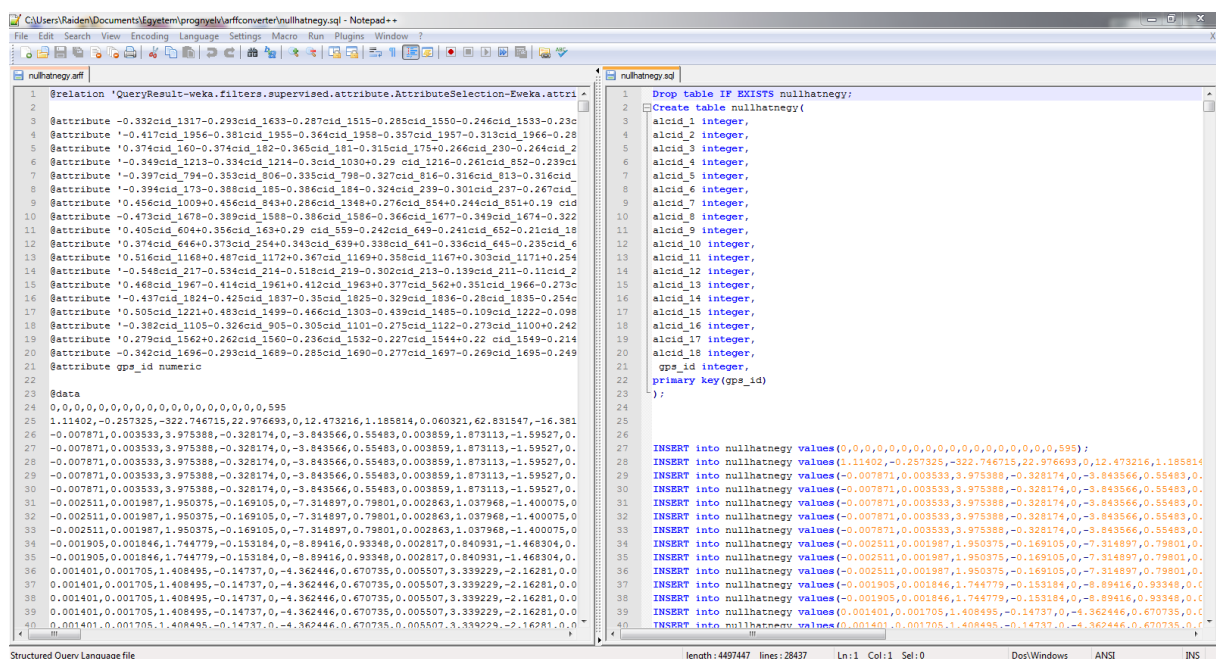
Ezekon az adatokon végeztem PCA tömörítési eljárást. A tömörítés során két értéket tudtam meghatározni:

- MaximumAttributeNames: maximális dimenzió szám, ennyi dimenziót engedek meg a tömörítés során
- varianceCovered: a kapott eredmény pontossága kezdethez mérve.



7. ábra Vágókép a WEKA programból

A kapott eredményeket elmentettem arff formátumban, mivel sajnos adatot visszatenni WEKA-n keresztül, sql-be menteni nem lehet. Ezért írtam egy java programot, ami egy arff fájlból, sql fájlt csinál. A program lényege, hogy soronként beolvassa az adatokat, lecseréli a sor elejét sql parancsokra.



8. ábra ARFF-SQL konvertálás

Miután sikerült átalakítanom „sql” formátumra, feltöltöttem az adatbázisba. Ezeket színeztam ki egy functions-al úgy, hogy a megegyező értékekkel kapott GPS pontoknak azonos színt adtam.

	gps_id integer	szin integer	alcid_1 integer	alcid_2 integer	alcid_3 integer	alcid_4 integer	alcid_5 integer	alcid_6 integer	alcid_7 integer	alcid_8 integer	alcid_9 integer	alcid_10 integer	alcid_11 integer	alcid_12 integer	alcid_13 integer	alcid_14 integer	alcid_15 integer	alcid_16 integer	alcid_17 integer
1	595	595	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2859	2859	1	0	-323	23	0	12	1	0	63	-16	0	0	0	0	0	0	0
3	3413	3413	0	0	4	0	0	-4	1	0	2	-2	0	0	0	0	0	0	0
4	3414	3413	0	0	4	0	0	-4	1	0	2	-2	0	0	0	0	0	0	0
5	3415	3413	0	0	4	0	0	-4	1	0	2	-2	0	0	0	0	0	0	0
6	3416	3413	0	0	4	0	0	-4	1	0	2	-2	0	0	0	0	0	0	0
7	3417	3413	0	0	4	0	0	-4	1	0	2	-2	0	0	0	0	0	0	0
8	3418	3418	0	0	2	0	0	-7	1	0	1	-1	0	0	0	0	0	0	0
9	3419	3418	0	0	2	0	0	-7	1	0	1	-1	0	0	0	0	0	0	0
10	3420	3418	0	0	2	0	0	-7	1	0	1	-1	0	0	0	0	0	0	0
11	3421	3421	0	0	2	0	0	-9	1	0	1	-1	0	0	0	0	0	0	0
12	3422	3421	0	0	2	0	0	-9	1	0	1	-1	0	0	0	0	0	0	0
13	3423	3423	0	0	1	0	0	-4	1	0	3	-2	0	0	0	0	0	0	0
14	3424	3423	0	0	1	0	0	-4	1	0	3	-2	0	0	0	0	0	0	0
15	3425	3423	0	0	1	0	0	-4	1	0	3	-2	0	0	0	0	0	0	0
16	3426	3423	0	0	1	0	0	-4	1	0	3	-2	0	0	0	0	0	0	0

9. ábra Színezett adatok táblája

4. Eredmények

A „Megvalósítás” c. fejezetben taglalt módon dolgoztam a félév során. Eredményeimet Microsoft Excel táblázatban prezentálom.

név	threshold	szín	Különbség	különbség az 1től
ot_colored	0.05	804	546	27606
nullahatketto_colored	0.062	1350	36	27060
nullhatnegy_colored	0.064	1386	110	27024
nullhathet_colored	0.067	1496	1573	26914
tiz_colored	0.1	3069	3801	25341
harminc_colored	0.3	6870	770	21540
negyven_colored	0.4	7640	550	20770
otven_colored	0.5	8190	163	20220
otvenotos_colored	0.55	8353	114	20057
hetven_colored	0.7	8467	68	19943
nyolcvan_colored	0.8	8535	19875	19875
Teljes	1	28410	0	0

1. táblázat PCA tömörítési eredmények

5. Következtetések

A félév végére nyert eredményeimet összevetve mérőtársam, Réti Dániel munkájával arra a következtetésre jutottam, hogy a PCA klasszifikációs eljárás hasznosnak bizonyult, mivel Réti Dániel „6_2-es szűrő” vizsgálata és az én 55%-os pontossággal futatott PCA eredményem esetén a 7 bázisállomást látó 224 mérési pontok esetén a szűrő 194 darab helyét határozta meg pontosan, míg a PCA 146 pont helyét. Viszont ami nem elhanyagolható, hogy a kettő összesen 146 pontban megegyezik.

6. Köszönetnyilvánítás

Köszönetemet szeretném kifejezni első sorban Tihanyi Attilának, konzulensemnek, aki készséggel fogadott önálló laboratóriumi munkára és bármikor állt rendelkezésemre, ha kérdésem volt. Külön köszönetet szeretnék mondani Dr. Lukács Gergelynek, aki rengeteget foglalkozott velem, segítségemre volt mind személyesen, mind otthon e-mail-ben. Tanácsokkal látott el a munkám során, illetve a dolgozatom készítésére és a kiselőadásomra vonatkozóan egyaránt. Köszönet továbbá mérőtársaimnak, akikkel sokat segítettük egymásnak munkánk során, igazi csapatot alkotva. Köszönöm!

7. Irodalomjegyzék

[Jeszenszky, 2011] <http://www.inf.unideb.hu/~jeszy/download/datamine/doc/pca.pdf>
2012.03.20. 12:20

[Bánhalmi, 2010] www.inf.u-szeged.hu/~banhalmi/adatb/2_eloadas.ppt 2012. 03.12. 15:30

[WEKA, 2010] <http://sourceforge.net/projects/weka/> 2012.03.25. 09:15

[Software, 2012] <http://www.cs.waikato.ac.nz/ml/weka/> 2012.03.10. 19:45

[WIKI, 2010] <http://hu.wikipedia.org/wiki/PostgreSQL> 2012.04.05. 13:40

[HUPWiki, 2012] <http://wiki.hup.hu/index.php/PostgreSQL> 2012.04.10. 15:20

[Majer, 2010] http://digitus.itk.ppke.hu/~tihanyia/HelyUltra/Majer_Julianna_Dipl.pdf
[2012.04.14. 20:15](#)

[WEKA, 2012] <http://weka.wikispaces.com/Databases#Setup-Driver> 2012.04.21. 22:45

8. Összefoglalás

A félév során az önálló laboratórium 1 tantárgy keretein belül végzett munkám első lépése az irodalomkutatás megfelelő módszerének elsajátítása volt. A keresési kulcsszavak megfelelő használata, a legnépszerűbb internetes kereső szervert használtam a Google-t. Csakis a megfelelő kulcsszavakkal lehet a rengeteg ömlesztett találatból kiszűrni a nekünk megfelelőt.

Feladatomból volt továbbá a metodikai betanulás, ezen belül is az adatgyűjtés az adott témában, adatértelmezés, ezen ismeretek megfelelő felhasználása a gyakorlati munkám során, illetve a precíz dokumentálás. Irodalomkutatásom során, internetes források alapján megismertem a helyes hivatkozás, utalás, idézés és a lábjegyzet készítés formai követelményeit.

Következő lépésként megismerkedtem a PCA keresési algoritmussal. Majd ezt követte az adatbányászati program, a WEKA letöltése és funkcióinak megismerése is. Utolsó, ámde nem elhanyagolható lépés, a PostgreSQL adatbázis nyelv megismerése volt. Ezután az következett, hogy elvégeztem a rám kiszabott méréseket, vizsgálatokat.

Önálló laboratóriumi munkám célkitűzése az volt, hogy önálló vizsgálatokkal, mérésekkel alátámasztva keressek egy optimális adatbázis-tömörítési eljárást, amellyel egyszerűbben tudunk dolgozni az adatbázisokon és jobban tudjuk reprezentálni a vizsgálni kívánt számszerűsített adatokat. A tömörítés lényege az, hogy az adatbázisban a kívánt adatok a lehető legtömörebbek, leg-helytakarékosabbak legyenek és a lehető legkönnyebben lehessen velük dolgozni.

Ezzel a módszerrel időt és helyet takarítunk meg, ezzel könnyebbé téve az adatok kezelését, az adatbázis jobb megismerhetőségét, összességében az adatbázison dolgozó szakember munkáját. A félév során a Pázmány Péter Katolikus Egyetem, Információs Technológiai Karának GSM adatbázisán végeztem klasszifikációs eljárást. Fő feladatomból az volt, hogy teszteljem a PCA tömörítési eljárást az egyetem GSM adatbázis szerverén. A félév célja, a PCA eljárás, a WEKA adatbányászati program és a PostgreSQL nyelv megismerése, jövőben elkészítendő diplomamunkám tudományos igényességű előkészítése volt, hogy felkészítsen a következő önálló feladatomra, ami maga a diploma dolgozat elkészítése. Ehhez ez a félév megadott minden ismeretet és előképzettséget. Önálló laboratóriumi munkámat koordinálta és felügyelte Tihanyi Attila.