

MNB-BME Együtműködés
2020/2021

Digitalizáció, mesterséges intelligencia és
adatkorszak Műhely



Az okostérkép fejlesztésének konceptuális alapjai, az adatbázis
sémája, az alkalmazással kapcsolatos funkcionális elvárások

Készítette:

Dr. Szakadát István

10.1.2.6. Okosváros és elektromobilitás

Budapest, 2021

Tartalomjegyzék

1. BEVEZETÉS	3
EGYSÉGES ADATSÉMA	5
ADATKÉSZLETEZÉS, ELŐSZÚRÉS	5
ELŐSZÚRT ADATCSOMAGOK	6
ALKALMAZÁS	6
2. ÁLTALÁNOS MODELLEZÉSI MEGFONTOLÁSOK, ELVÁRÁSOK	7
3. A FÖLDRAJZI ÉS SZEMANTIKAI TÉR METSZETÉBEN	7
ADATSÉMA	7
<i>Geoontológia felépítése</i>	<i>7</i>
<i>elkülönített nyelvi modell</i>	<i>8</i>
<i>időkezelés eseményeken keresztül</i>	<i>8</i>
<i>geoentitások megkötése köznévtérben: geoentitások tipizálása</i>	<i>8</i>
<i>geoentitások megkötése a tulajdonnévtérben: geoentitások megnevezései</i>	<i>9</i>
<i>Geoentitások megkötése a földrajzi térben: geoentitások geometriája</i>	<i>9</i>
<i>Geoentitások közti partitív kapcsolatok</i>	<i>10</i>
<i>Geoentitások kapcsolatai más entitásokkal</i>	<i>10</i>
STATISZTIKAI ADATOK, KEMÉNY ADATRÉTEG	11
<i>Adatbázis-leírások</i>	<i>11</i>
A STATISZTIKAI ADATOK KÖRÉNEK BŐVÍTÉSE	13
A TELEPÜLÉSEK INTÉZMÉNYI-SZOLGÁLTATÁSI KÍNÁLATA	13
<i>Indikátorok</i>	<i>14</i>
DASHBOARD FEJLESZTÉS	15
4. DEFINÍCIÓK	15
POLICY-KÉRDÉSEK	16

Az okostérkép fejlesztésének konceptuális alapjai, az adatbázis sémája, az alkalmazással kapcsolatos funkcionális elvárások

1. Bevezetés

A földrajzi térhez köthető adatokat régóta kezeljük számítógép segítségével. A térinformatikai eszközöket, megoldásokat évtizedek óta használunk, és az okosváros témakörében felmerülő ötletek, projektek mind igénylik, hasznosítják a térinformatika által kínált megoldásokat. Mivel az egyetemen fokozódó érdeklődés irányul a magyar települési rendszer fejlődésére, a települések térbeliséghez köthető problémáinak elemzésére, célszerű volt az okosváros fejlesztésekkel kapcsolatos víziókat, projektelképzeléseket bővíteni, és a városokra irányuló fókuszot kitágítani a településekre általában. A térhez köthető adatkörből rengeteg van, és ezeket a térinformatikai eszközök segítségével reprezentálni, elemezni lehet. Az adatvezérelt korszakban egyre több adat áll rendelkezésre, és ezeket a településekhez (földrajzi térhez) kötve egyre többet tudhatunk meg adott földrajzi egységről, ha az egyre bővülő adatrétegekre támaszkodva megfelelő elemzéseket végzünk. Sok olyan probléma van, amikor az elemzések akkor lehetnek sikeresek, megalapozottak, ha minél több adatkört veszünk egyszerre figyelembe. Ehhez folyamatosan növelni kell a különböző forrásokból, különböző platformokról, különböző időből származó adatok integráltságát. Ez az integráltságnövelés megfelelő módszerekkel, az egységesítést megalapozó szabványokkal és rengeteg munkával elérhető. Minél több adatkör, minél több idősor kerül be egy rendszerbe, annál több és pontosabb elemzési lehetőség adódik a rendszer használói számára. A cél az, hogy egyetlen, integrált térinformatikai alapon működő téradatrendszert dolgozzunk ki, tartsunk fent és fejlesszünk tovább, amelynek megfelelő alapokat szolgáltat a mindenkor elemzések számára.

A BME-n sok karon van a földrajzi térrel kapcsolatos adatgyűjtés, illetve térbeli adatelemzési kompetencia: a Közlekedési kar forgalomszervezési kérdésekkel foglalkozik, az Építőmérnöki karon időjárás, természetföldrajzi, vízrajzi stb. adatokat elemez, a Gépészkaron energetikai rendszerek térbeli problémáit elemzik, a villamosenergia hálózatok fejlesztésével vagy az okosvárosokkal kapcsolatos kérdéseket vizsgálják, a GTK-n társadalomföldrajzi, társadalomstatisztikai, politikai elemzések térbeli meghatározottságát kutatják.

A Smart Map projekt elindításával az a cél, hogy ezeket a kari kompetenciákat, elkülönült adatbázisokat egységesítsük, egy platformra tereljük, és kihasználjuk az így keletkező szinergikus lehetőségeket. Az integrált adatbázis kialakításának, a közösen végzendő elemzési munka elindításának feltétele, hogy a rendelkezésre álló adatok, adatkörök átláthatók, könnyen megtalálhatóak, az elemzésekbe gyorsan bevonhatók legyenek. Annak érdekében, hogy a sokszempontú értékelésekhez, elemzésekhez szükséges adatköröket az elemzések során be lehessen tölteni a rendszerbe, olyan dashboard-jellegű alkalmazást készítettünk, amely biztosítja adott szempontú adatkörök, indikátorok, elemzési technikák előzetes kiválasztását, majd ezen elemek térinformatikai reprezentációját.

A téradatrendszer adatai csak akkor használhatók igazán elemzésekhez, ha a kutatók gyorsan és könnyen hozzáférhetnek az adatokhoz. Ez az elvárás egyre nehezebben teljesíthető, ahogy egyre több adatkörrel bővül a rendszer egésze. A hozzáférés problémájának megoldására van szükség egy olyan alkalmazásra, amely segíti a rendszerben tárolt adatok gyors megismerését, kikereshetőségét.

Ezt a feladatot látja el az okostérkép alkalmazás. Ezt az alkalmazást egyfajta dashboardként lehet értelmezni, ami képes az előre meghatározott adatköröket a térhez kapcsolva bemutatni. A rendszer egyik legfontosabb funkciója az, hogy a digitális térképen barangolva, a térképen elhelyezett fókuszponthoz köthető adatokat folyamatosan megjeleníti a képernyő kijelölt felületrészében. Ez a funkcionalitás lokatív média sajátossága, amire a mobil applikációk egy része is képes, amikor a mobiltelefonnal a valós földrajzi térben mozogva az alkalmazás mindig új, a térhez kötött adatokat kínál fel a felhasználó számára. Az újdonság az okostérkép alkalmazásban az, hogy jelen esetben a virtuális (digitális) térben kell/lehet mozogni, és e virtuális mozgás szerint változik folyamatosan a megjelenített tartalom.

Az alkalmazásban meg lehet tekinteni a rendelkezésre álló geoentitások térbeli elhelyezkedését. Ez a puszta reprezentációs funkció arra jó, hogy egy elemzések számára konkrét adatköröket lehessen előzetesen megtekinteni és kiválogatni a rendszer által felkínált adatrétegekből. Az alkalmazásban többféle, előre legyártott adatgyűjtemény szerint lehet barangolni a digitális földrajzi térben. Egyelőre kézzel lehet beállítani az ilyen barangolási tematikák tartalmát. Egy következő fejlesztési menetben lehet egy olyan felületet (egy metadashboard-ot) elkészíteni, ahol könnyen lehet új tematikákat összeállítani. Amikor egy adott tematika mentén barangolva az éppen mutatott településre kíváncsi az ember, akkor egy kattintással lehet elugrani egy olyan oldalra, ahol sokkal több adat lesz elérhető az adott településre vonatkozóan. A rendszer módot ad kiválasztott települések adatainak összehasonlítására is.

Az okostérkép alkalmazás egyelőre ideiglenes webcímen érhető el (radar.o-o.hu). A szükséges egyetemi szintű egyeztetések után egyetemi webcím alá fogjuk áttenni az alkalmazást.

Az okostérkép alkalmazás akkor lehet sikeres, ha valóban egy nagyon magasan integrált, egyszerű és egyértelmű adatsémára támaszkodó, egységes adatbázisból lehet kinyerni az adatokat. Ehhez a fejlesztés előtt ki kellett alakítani azt a környezetet, azt az adatbázissémát, azt a szempontrendszert, amely biztosítani képes a hosszútávú építkezést. Olyan megoldásokat kellett választani, amelyek a gépi adatfeldolgozás szempontjait is figyelembe veszik. Az egyértelműség érdekében a földrajzi egység (geoentitás) számára önálló – földrajzi – névteret hoztunk létre, amelynek segítségével minden egyértelműen azonosíthatjuk a térben is időben folyamatosan változó (keletkező, megszűnő) földrajzi egységeket. A rendszer adatköreinek jelentős része másfajta entitások (személyek, szervezetek, intézmények stb.) és a földrajzi tér közti kapcsolatokat írja le, ami egységes szerkezetű adatsémával lekezelhető. A konceptuális kérdéseket, a választott megoldásokat egy tanulmányban írtuk le.

Az okostérkép fontos komponenseit jelentik az előzetesen, szakértői munkával összeállított adatgyűjtemények. Ezek alapozzák meg azt a funkciót, hogy a rengeteg rendelkezésre álló adat közül azok és csak azok az adatkörök jelenjenek meg, amelyek valamely érdeklődés, elemzés számára fontosak. Az okostérkép ugyanabban a földrajzi térben ajánl adatbarangolást, de a más és más tematika mentén megjeleníthető adatgyűjtemények más és más felhasználói igényt képesek kiszolgálni. A projekt során többféle tematikát (adatgyűjteményt) dolgoztunk ki (fenntarthatósági adatok, településértékelési rendszer és TSTAR kapcsolata, zsugorodó települések mutatói), amelyekről kisebb rész tanulmányok születtek.

A közelmúltban elindult egy projekt, amelynek keretében egy olyan – országos szintű adatforrásokra támaszkodó – adatrendszer kezdtünk el felépíteni, amely minden településre vonatkozóan tartalmaz adatokat, és így tetszőleges települést be lehet vonni az elemzési folyamatba. Ezzel az adatgyűjtési fókusszal természetesen veszünk az elemzések pontosságán, hiszen nem mindig állnak rendelkezésre elég finom felbontású adatok az országos szintű adatrendszerekben, viszont így hatalmas adatmennyiséghez férhetünk hozzá, ami lehetővé teszi új – MI alapú – elemzési technológiák alkalmazását is.

Az országos szintű településértékelési rendszer adatbázisába első lépésként beépítettük a KSH által évente kiadott TSTAR társadalomstatisztikai adatrendszer éves adatállományait (1990-től

kezdődően). Ezt úgy valósítottuk meg, hogy a főkategóriák alá hierarchikus módon besoroltuk a KSH-változókat, a települési, földrajzi adatok kezelését pedig egységes földrajzi névtérre támaszkodva valósítottuk meg. A földrajzi névtéren belül elkezdtünk külön rétegbe építeni olyan intézményi adatokat, mint vasútállomások, benzinkutak, iskolák, községházák, templomok, gyógyszertárak, orvosi rendelők, művelődési házak, posták stb.

Ezzel a koncepcióváltással rengeteg települési adathoz lehet hozzájutni, viszont újfajta hozzáférési probléma is keletkezett: a rendelkezésre álló túl sok adat nehezítette az éppen szükséges adatok megtalálását. Ezért terveztünk meg és fejlesztettünk ki egy olyan felületnek a prototípusát, amely a rendelkezésre álló adatkörök megtalálásában nyújthat segítséget. A prototípus ideiglenes címe: radar.o-o-o.hu. A térinformatikai alkalmazásban több olyan dashboard-szerű felületet lehet igénybe venni, amelyek mind előre megadott, de szabadon választható témák adatait jelenítik meg a földrajzi térhez kapcsolódóan. A prototípusban elkészítettünk több, társadalomstatisztikai mutatókat megjelenítő felületet, de van olyan oldal is, ahol természetföldrajzi mutatókat mutat a rendszer aszerint, hogy éppen hol jár az ember a települési térben. Van olyan oldal is, ahol egy településre sokkal több (elvileg 3 ezer) adattípus adatait lehet megtekinteni, és van olyan oldal is, ahol a felhasználó által egyszerre több települést kiválasztva a települések adatainak összehasonlítására van lehetőség.

A következő lépésekben az így kialakított települési adatrendszerbe további adatrétegeket érdemes betölteni, hogy minél több – egységes szerketű, egységes névtérre vetített adat álljon rendelkezésre. A közös adatrendszerbe nem csak az adatokat, illetve az ezekre támaszkodó összetett mutatókat, indikátorokat, de az elemzéshez szükséges műveleteket, algoritmusokat is be lehet integrálni, és az így felépítendő rendszert úgy lehet értelmezni, használni, hogy a jövőbeni elemzésekhez ebből a rendszerből lehet adatokat, indikátorokat, műveleteket, algoritmusokat leválogatni.

Egységes adatséma

A rendszer számára olyan adatsémát alakítottunk ki, amely tetszőleges forrásból származó tetszőleges adatot fogadni képes. Az adatséma központi magját egy földrajzi névtér jelenti, amely az összes geoentitás alapadatait tartalmazza úgy, hogy a tényleges földrajzi helyeket (geoentitás-példányokat) a széles körben használt háromféle geoentitástípus alá vannak rendelve. A földrajzi helyek vagy más típusú entitások (például személyek, testületek, dokumentumok stb.) közti kapcsolatok kezelésére a linked data világ triplet-logikájához közel álló, de azért attól eltérő megoldást választottunk.

Adatkészletezés, előszűrés

A kialakított rendszer egyik nagy előnye, hogy a sokféle forrásból, sokféle időből származó adatkörök egységes, integrált rendszerben vannak kezelve. Ennek az előnyös helyzetnek van egy hátulütője is, mégpedig az, hogy a túl sok adat együttes „jelenléte” megnehezíti a konkrét reprezentációhoz, konkrét elemzéshez éppen szükséges adatok megtalálását.

A konkrét tér- és időbeli elemzésekhez, a konkrét tér- és időbeli adatok megjelenítéséhez tehát szükség van egy olyan „szolgáltatásra”, felületre, ahol ki lehet válogatni az aktuálisan felhasználni kívánt adatokat. Ez egyfajta szűrést jelent.

Előszűrt adatcsomagok

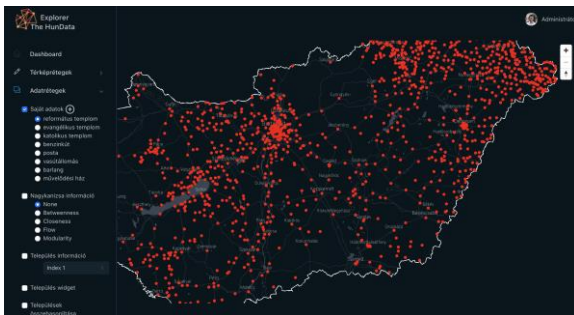
- okosváros értékelési tematika
- környezetvédelmi, fenntarthatósági tematika
- zsugorodó város tematika

Alkalmazás

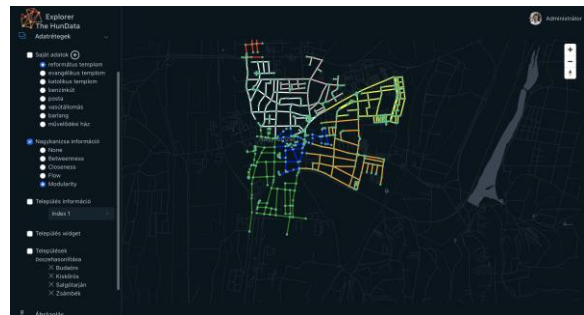
Az alkalmazás nem tartalmazza az összes rendelkezésre álló adatot, és csak néhány alapfunkciót implementáltunk.

Az alkalmazás ideiglenes webcíme:
radar.o-o.hu

Néhány képernyőkép az alkalmazás funkcionális oldalai közül.



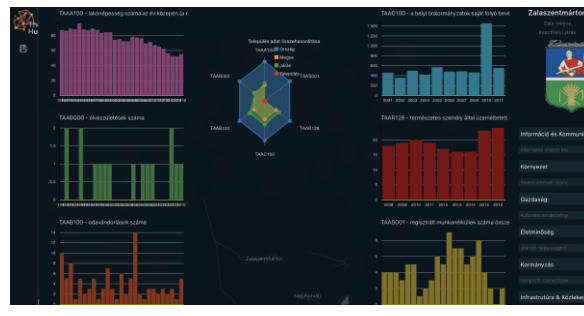
intézményi adatrétegek választása



települési utcaszerkezet absztrakt hálóstruktúrája



TSTAR-mutatók barangolásához



a kiválasztott település adatai főkategóriáinként



természeti mutatók barangolásához



négy település összehasonlító oldala

2. Általános modellezési megfontolások, elvárások

- egyetemi koalíció
 - közlekedésmérnöki kar: közlekedésforgalmi adatok, közlekedés infrastruktúra
 - építészmérnöki kar: épületállomány, urbanisztika
 - építőmérnöki kar: természetföldrajzi adatok (barlangok, vízforrások, meteorológia stb.)
 - gépészmérnöki kar: épületek energetikai adatai (okosépületek)
 - gazdaság- és társadalomtudományi kar: társadalomstatistika, választás, fenntarthatóság, önkormányzati költségvetés
 - természettudományi kar: elemzés
- közös adatplatform: BME Radar (mintapélda: Stanford Dataverse)
 - adatbázis + műveletbázis (adatkörök + függvénykönyvtárak, algoritmusok)
 - szinergia: OSM földrajzi tér + szemantikai tér
 - kinek, mit szolgáltat az egyetem?
- elemzés
 - MI-algoritmusok alkalmazása a nagytömegű települési adatokra
- reprezentáció
 - felfedezés: hipotézisgenerátor
 - települési önreprezentáció, önreflexió támogatása

3. A földrajzi és szemantikai tér metszetében

A földrajzi térhez köthető adatokat a nyelvi-szemantikai tér segítségével tudjuk értelmezni, és az adatmodellnek ezt a kettősséget jól kell tudnia kezelnie.

Adatséma

A rendszer számára olyan adatsémát alakítottunk ki, amely tetszőleges forrásból származó tetszőleges adatot fogadni képes. Az adatséma központi magját egy földrajzi névtér jelenti, amely az összes geoentitás alapadatait tartalmazza úgy, hogy a tényleges földrajzi helyeket (geoentitás-példányokat) a széles körben használt háromféle geoentitástípus alá vannak rendelve (pontszerű, vonalszerű és poligonszerű entitások).

Geoontológia felépítése

A fogalmi tér elemeivel írhatjuk le a földrajzi adatok szemantikus tartalmát. Ezt a világtudást geoontológiában tárolhatjuk, amelynek általános részét kifejezhetjük egyszerű tripletek segítségével.

concept(id)	relation(id)	otherconcept(id)
főváros	is-a-type-of	város

város	is-a-type-of	település
település	is-a-type-of	poligonszerű geoentitás
település	is-part-of	megye
megye	is-part-of	ország

elkülönített nyelvi modell

A tulajdonneveket, közneveket külön nyelvi modellben kezeljük. Minden nyelvi konstrukciónak saját azonosítója van és ezekre lehet hivatkozni a geo-modellben.

időkezelés eseményeken keresztül

Az eseménynek két fő típusát elkülönítve mondhatjuk, hogy vannak pontszerű és folyamatszerű események, előbbiekhez egyetlen időpontot rendelhetünk, utóbbiakhoz két időpontot (az esemény kezdetét és végét) rendelhetjük.

A folyamatszerű esemény kompozit esemény, hiszen több eseményből áll.

Az események időbeliségével kapcsolatos jól ismert nehézség az események időpontosságának/időpontatlanságának kérdése. Vannak események, amelyeknek másodpercre, napra, hónapra, évre, évtizedre stb. pontosan tudjuk az idejét, és ilyen körülmények között nem lehet egységes adatszerkezetet használni az eltérő pontosságú események leírásához: egy évtized-pontosságú eseményt nem lehet dátum-adattípussal leírni. A probléma megoldása az lehet, hogy minden esemény idejét dátum-típusúként vesszük fel, és egy új adattípussal, az időpontossággal jelezzük azt, hogy az adott dátumból mennyi az „értékes” érték.

event(id)	eventtype(id)	eventdate	timeprecision(id)
2221	megnevezés kezdete	1983-1-1	év-pontos
3259	megnevezés vége	1998-12-31	nap-pontos
98415	megnevezés kezdete	1400-1-1	évszázad-pontos

geoentitások megkötése köznévtérben: geoentitások tipizálása

Egy földrajzi helyet (geoentitást) mindig el kell helyoznünk a szimbolikus térben, meg kell adni a típusát, hogy példányként milyen fogalom alá sorolható. Ehhez természetesen szükség van a földrajzi helyek ontológiájára, amelynek segítségével képesek lehetünk minden földrajzi helyet megragadni.

A

geo(id)	relation(id)	type(id)	event(id)
Balaton	is-instance-of	község	1948-04-01 (8372)
Sztálinváros	is-instance-of	város	1951 (48372)
Budapest	is-instance-of	főváros	1873 (2123928)

geoentitások megkötése a tulajdonnévtérben: geoentitások megnevezései

a könnyebb hivatkozás, azonosíthatóság kedvéért a földrajzi entitásoknak nevet vagy neveket adhatunk

a név valamilyen szimbolikus entitás, aminek nincs természetes időbelisége, a megnevezés valamilyen esemény, aminek van természetes időbelisége

a geoentitásokhoz nem neveket, hanem megnevezéseket rendelünk

geoinstance nametype event name

geo(id)	nametype(id)	name(id)	event(id)
Budapest XI. ker. (7234)	hivatalos (település)név	Budapest XI. kerület (43824)	1950 (219128)
Budapest XI. ker. (7234)	közösségi név	Újbuda (89472)	1990-es évek (43911)
Szeged (312)	hivatalos (település)név	Szeged (32710)	1300-as évek (17173)

Geoentitások megkötése a földrajzi térben: geoentitások geometriája

Helypontosság

Minden geoentitás felfogható a földrajzi tér valamilyen szegmenseként, tehát elvileg mindig meg kell tudnunk adni valamely geoentitás földrajzi elhelyezkedését (geometriáját), de a gyakorlatban ez nem mindig teljesül. Előfordulhat, hogy nem vagyunk képesek kellő pontossággal megadni egy geoentitás földrajzi elhelyezkedését.

Egy geometria több geoentitáshoz is hozzákapcsolható, tehát egy geometria nem jelöl meg egyértelműen egy geoentitást. A TSTAR-tematikában a település és a poligonok között 1:1-es kapcsolat van, de egy települési poligon megegyezhet a választási adatkör egy szavazókörének poligonjával. A földrajzi tér valamilyen szegmensét csak egy azonosítóval láthatjuk el azon túl, hogy megadjuk a pontos koordinátáit.

geometria geocode

geometry(id)	type(id)	geometry
813	poligon	0106000020E61000000100000001030000000100000022000000833C1DA8B
115	pont	41ACEA3240EE298FB96BA447403E8013D347EB3240E60CE24E22A3474065D

geoinstance geometria event

geo(id)	datatype(id)	geometry(id)	event(id)
7234	has-polygon	813	1950-03-15
7234	has-polygon	115	1998-01-01

amennyiben egy geoentitásnak ismerjük a földrajzi koordinátáit, akkor minden más geoentitáshoz való kapcsolatát származtatni tudjuk a geoentitások geometriai adatai alapján, de egyrészt, ameddig nincs meg minden geoentitás pontos geometriája, addig nem lehet csak számolásokra támaszkodni, másrészt a számítási feladatok csökkentése érdekében érdemes redundáns módon is kifejezni a geoentitások közti partitív viszonyokat.

Geoentitások közti partitív kapcsolatok

geoinstance IS-A-Partitive-Subordinate-Of geoinstance2 event

geo(id)	datatype(id)	othergeoid(id)	event(id)
Pécs (7234)	is-part-of	Baranya megye (813)	1950-01-01

Geoentitások kapcsolatai más entitásokkal

A földrajzi helyek vagy más típusú entitások (például személyek, testületek, dokumentumok stb.) közti kapcsolatok kezelésére a linked data világ triplet-logikájához közel álló, de azért attól eltérő megoldást választottunk.

database	agent(id)	datatype(id)	geo(id)	event(id)	value
TSTAR		lakónépesség száma	Szeged (312)	1990	546
választás	Fidesz	pártlistás szavazata	Vékény (9313)	2014	234
választás	Kupa Mihály	egyéni jelölt szavazata	Vatta (5114)	2002	13
NAV 1%		felajánlott pénz	Vaszar (2456)	2013	385100

mértékegységek

datatype(id)	relation(id)	concept(id)
felajánlott pénz	measure_unit	forint
lakónépesség száma	measure_unit	fő
pártlistás szavazat	measure_unit	szavazatszám

ágensek közti kapcsolatok

person(id)	relation(id)	corporate(id)	event(id)
Kupa Mihály	egyéni képviselőjelölt	Centrum Párt	2002

Statisztikai adatok, kemény adatréteg

Rendelkezésünkre áll egy idősoros, földrajzi térhez kapcsolt, társadalomstatistikai adatbázis. Az adatbázis két fő szegmensét adják az 1990 óta tartott parlamenti választások, illetve a KSH által évente kiadott TSTAR (településstatistikai adatrendszer) adatai, 1990 óta. Az választási adatok szavazóköri szinten állnak rendelkezésre (alkalmanként 10 ezres szavazókörrel), de innen nyilván lehet települési szintre aggregálni az adatokat. A TSTAR adatok települési szintre vonatkoznak (alkalmanként 3200-3300 településsel). Települési szinten minden adat a földrajzi térhez köthető, az adatbázis pontosan követi 1990-től fogva a települések (területi, közigazgatási, névbeli) változásait. Ez az adatszegmens jelenti az adatrendszer kemény adatrétegét, amelyben településstatistikai adatok állnak rendelkezésre.

Az alprojekt célja a rendelkezésre álló adatok sokszempontú elemzése mesterséges intelligencia eszközök segítségével. Ehhez természetesen szükséges lehet igénybe venni statisztikai és más adattudományi technológiákat is. A praktikus cél itt az, hogy minél többféle módon lekérdezhetővé, bejárhatóvá tegyük az adatbázist. Meg lehet tervezni egy olyan okos dashboard-ot, ami az adatrendszerre ültetve a laikus felhasználók számára is könnyen hozzáférhetővé teszi az adatokat, és sok okosságot ki lehet vele nyerni a rendszerből.

Érdeemes a projekt két eredményét, az egységes településértékelési rendszert, illetve az intelligens adatrendszert összekapcsolni egymással. Ennek az lehet a módja, hogy az értékelési keretrendszert összekapcsoljuk az adatrendszerünk szemantikai terével. Ehhez arra van szükség, hogy egységes keretrendszer kialakítása során figyelembe vegyük azt a szemantikai teret, amelyet fel kell építeni az települési adatrendszer adattípusainak szemantikai tartalma alapján. Az értékelési keretrendszeren belül ezáltal lehetőség nyílik arra, hogy bizonyos értékelési szempontokat kiválasztva adatokat lehessen kérni az adatrendszerből, amelyeket aztán hasznosítani lehet az értékelési folyamat során.

Adatbázis-leírások

Rendelkezésre álló adatbázisok

- népszámlálási adatok
 - 1990
 - 85 ezer számlálóbiztosítási körzet (Budapestre 15 ezer poligon: a legfinomabb felbontású térkép forever)
 - beszerezhetőek lennének a 2001-es és 2011-es népszámlálás adatai is
- KSH TSTAR (településstatistikai) adatok
 - 1990-2018, évente, 3200 település
 - település szinten változáskezelte poligonok (OSM-kompatibilis)
 - ~2500 adattípus (az adattípusok leírása (fagráfba rendezve): <https://o-o-o.hu/tstar/metadata.php>)
 - minta térképes megjelenítésre: <https://o-o-o.hu/hits.php?hitstypeid=4>
 - az adatok „triplet”-sémába rendezve, összesen ~45 millió rekord, évenként 1-3 millió rekord
- parlamenti választási adatok
 - 1990-2018, 8 választás
 - 10 ezer szavazókör, országosan település szinten változáskezelte poligonok
 - Budapestre és pár városra szavazókör szinten poligonok, 1990-1998, megszerezhető a teljes országos szavazóköri poligonhálózat a 2010-es évekre

- választásonként ~2-3-szor 100-150 ezer rekord
- térképes adatmegjelenítés: <https://o-o-o.hu/pol/map.php>
- önkormányzati költségvetési adatok 2010-es években
 - GTK Pénzügy Tanszék
 - az adatok excel-táblában vannak, a KSH-kódok alapján könnyen map-pelhetők a település szintű poligonokra
- NAV 1%-os alapítványi felajánlások adatai
 - 2003-2020 (adatok közterület/utca szinten)
 - OSM-utcanévjegyzékére részben map-pelve
 - évente ~20-25 ezer rekord
- KSH átlagos ingatlanár adatok
 - 1997-2019
 - közterület/utca szinten, évente ~5000
 - OSM-utcanévjegyzékére részben map-pelve
- Oszkár forgalomstatistikai adatok
 - car sharing szolgáltatás (települések mint kiinduló pontok, illetve célpontok)
 - Körny.gazd: Tanszék + SzocKom Tanszék

sémaleírások

Az adatbázisok sémái természetesen különböznek, de átalakíthatók egy közös szerkezetbe:
hely | idő | adattípus | adatérték

Kezdetben a két legnagyobb méretű adatbázis elemzésével érdemes elindulni.

TSTAR + 1990-es népszámlálás

adattípus	leírás
datatype/datatypeid	adattípus neve, azonosítója
geoname/geonameid (geoid)	a település neve, azonosítója; egyértelmű kapcsolat a geoid-vel (legkisebb területi egység: település; aggregációs lépcső: település, járás, megye, ország)
year	az adat időbelisége
data/value	az adattípus tényleges értéke az adott településen

parlamenti választás

adattípus	leírás
event/eventid + eventtype/eventtypeid	választási esemény neve, dátuma + esemény típusa (pártlistás vagy egyéni szavazat leadása, illetve első vagy második fordulás választás)
geoname/geonameid (geoid)	a földrajzi hely neve, azonosítója; egyértelmű kapcsolat a geoid-vel; (legkisebb területi egység: szavazókör; aggregációs lépcső: szavazókör, település, járás, választókerület, megye, ország)
datatype/candidate	pártlista vagy egyéni jelölt neve, azonosítója
data/vote	adatérték: a földrajzi helyen a jelöltre leadott szavazatok száma

A statisztikai adatok körének bővítése

A projekt során érdemes a más forrásokból származó, természetföldrajzi, időjárési és egyéb adatokat is integrálni az adatrendszerbe. Ennek az integrációs munkának lehet egy olyan összetevője is, amely azt méri fel és azt készíti elő, hogy hogyan lehet további potenciális adatköröket integrálni egy települési adatrendszer keretei közé, ennek milyen technikai, adatgyűjtési, szabványosítási stb. feladatai vannak. E tervezési munka során különös hangsúlyt kell adni a térhez köthető, térinformatikai jellegű adatok integrálásával kapcsolatos feladatoknak.

A települések intézményi-szolgáltatási kínálata

A mindenkori településszerkezet fenntarthatóságának, a települések élhetőségének leírásához, értelmezéséhez nem elég „kemény”, statisztikai adatokat gyűjteni, kezelni, de szükség van arra, hogy a településekre puha adatokat gyűjtsünk és azok alapján próbáljunk meg elemezni, értelmezni, előrejelezni települési folyamatokat. Ehhez minél több "puha", intézményi, kulturális jellegű adat begyűjtésére van szükség a településekre vonatkozóan. A projekt elején összegyűjtjük és az integrált rendszer egy rétegébe építjük a települések közösségi, kulturális (georeferált) adatait (iskola, könyvtár, óvoda, múzeum, posta, gyógyszertár, templom, városháza, művelődési ház, polgármesteri hivatal, vasútállomás, kórház, rendelő, sportlétesítmény, benzinkút stb.) Ez nem mind kulturális intézmény, nem mind puha adat, de ezek együtteséből, illetve ezek térbeli viszonyaiból különféle mutatókat lehet képezni arra, hogy milyen intézményi ellátottsága van a településnek, milyen szolgáltatásokhoz, milyen módon, milyen gyorsan, milyen „költségekkel” tud hozzájutni a település átlagpolgára. Itt lehet újrarahasznosítani a kemény adatréteg olyan adattípusait, mint az ingatlanárak alakulása vagy az alapítványi felajánlások pénzügyi nagyságrendje, illetve ezek változásai.

intézménytípus	lista	geometria
iskola	megvan	részben megvan
óvoda	megvan	részben megvan
egyetem, főiskola		
könyvtár	megvan	részben megvan
múzeum		
posta	megvan	részben megvan
templom (katolikus, református, evangélikus, ortodox, zsidó stb.)	megvan, ellenőrzés	részben megvan
városháza, községháza, polgármesteri hivatal, körjegyzőség	megvan	részben megvan
színház		
mozi		
művelődési ház, közösségi ház, teleház, faluház	részben megvan	
vasútállomás	megvan	megvan, ellenőrzés
buszállomás	megvan (3 megye hiányzik)	
benzinkút	megvan	megvan, ellenőrzés

kórház		
orvosi rendelő, egészségház, közzet i orvos		
gyógyszertár	megvan	részben megvan
sportlétesítmény		
sportpálya		https://www.magyarfutball.hu/hu/stadion/2913
alapítvány (1%-os)	megvan: 2003- 2020	részben megvan
bolt		
turisztikai nevezetesség (vár, várrom, kastély, park, fürdő)		
rendőrség		
bankfiók		

Indikátorok

települések térstrukturális jellegzetességei

- településközpont (lakóterületre számolva, teljes területre számolva, kézzel kijelölve)
- lakóterület (residential area) és teljes terület aránya
- úthálózat hossza (lakóterületen belül, teljes területen belül)
- úthálózat „bonyolultsága” (heterarchia/hierarchia „skálán”)
 - <https://www.mdpi.com/2071-1050/12/2/628/pdf>
 -
- belépési pontok száma: hány út vezet be a településre
- településrész szegmentálási fok (mennyire lehet „elkülönült” településrészekre tagolni a települést)
- utcák orientációja
 - <https://geoffboeing.com/2018/07/comparing-city-street-orientations/>
 - <https://www.bloomberg.com/news/articles/2014-03-26/is-your-city-s-street-grid-really-a-grid>
 - <https://www.bloomberg.com/news/articles/2018-07-26/mapping-the-hidden-patterns-of-cities>
 - <https://github.com/gboeing/osmnx-examples/blob/master/notebooks/17-street-network-orientations.ipynb>
 - <https://github.com/mourner/road-orientation-map>
 - <https://blog.mapbox.com/visualizing-street-orientations-on-an-interactive-map-1eefa6002afc>

fizikai elérhetőség

- autóbusz elérhetőség (hány buszmegálló van, milyen gyakran állnak meg buszok, hány településre lehet közvetlen módon eljutni)
- benzinkutak elérhetősége (milyen messze van a településközponttól a benzinkút)
- oszkar-elérhetőség (mennyi – be-, illetve kimutató – oszkar-utazás van)

- gépkocsi-ellátottság (tstar)
- vízi elérhetőség
- légi elérhetőség
- taxi-elérhetőség (milyen messze van a településközponttól a taxi-szolgáltatás)

társadalomstatistikai jellemzők (tstar)

- népsűrűség
- népmozgalmi jellemzők

szimbolikus elérhetőség

- postai elérhetőség (milyen messze van a településközponttól a posta)
- kulturális POI-k térstruktúrája

Dashboard fejlesztés

Amikor rengeteg – térhez és időhöz köthető – adat áll rendelkezésre, akkor fontos kérdés az, hogy a döntéshozó honnan tudja, hogy milyen adatok állnak rendelkezésre és hogyan találja meg az őt érdeklő adattípusok konkrét értékeit. Ezt a fajta információs igényt úgy értelmezhetjük, hogy

- az adattípusok valamilyen szimbolikus térbe (fogalmi struktúrába) rendezhetők,
- az adattípusok a földrajzi térhez köthetők
- az adattípusok (legalább részben) az időhöz köthetők (másként: időben változnak)

Mivel kétdimenziós reprezentációs felületeink vannak, arra van mód, hogy a földrajzi tér két dimenzióját kezeljük, de ebben az esetben nincs "helyünk" sem az idő, sem a szimbolikus-fogalmi tér kategóriáinak kezelésére. A kétdimenziós reprezentációs térbe úgy adhatunk át további információt, hogy színeket vagy ikonokat alkalmazunk adott jelentéssel

A fenti probléma megoldása a reprezentációs teret meg kell osztanunk és az adattér mellett olyan metateret kell elkülönítenünk, ahol módot adunk a felhasználónak, hogy keressen/szűrjön a szimbolikus térben, és a keresés/szűrés eredményét újra és újra megjeleníttethesse az adattérben.

4. Definíciók

kereső	olyan felület, ahol adattípusban és adatértékben együtt lehet keresni
metakereső	olyan felület, ahol az adattípusokban lehet keresni, hogy a következő lépésben adatértékhez juthassunk
metadashboard	olyan felület, ahol be lehet állítani, hogy adott dashboardon milyen adattípusok, hogyan jelenjenek meg
dashboard	olyan felület, ahol előzetesen kiválogatott adattípusok alapján automatikusan megjelennek az adatok az előre beállított paraméterek és szempontok szerint
térképes dashboard	olyan felület, amely az adatokat – alapértelmezés szerint – a földrajzi térhez kapcsolódóan jeleníti meg úgy, hogy a műszerfalat egy térképes alkalmazás felületére helyezik el. Idősoros megjelenítéshez ez nem igazán alkalmas.
grafikonos dashboard	olyan felület, amely az adatokat grafikonos (vagy táblázatos) formában jeleníti meg. Idősoros elemzésekhez ez alkalmas.

Policy-kérdések

nagytömegű, szerteágazó, sokforrású adatvagyonnal kapcsolatban

- terminológiai tisztázás, információtipológia
 - adat, adatbázis hol helyezhető el (vizuális, auditív + nyelvi, nem nyelvi)
 - állami, közösségi, privát adat – közadat
 - társadalmi (konstruált) vs. természeti (mért) adat
- mi az érték az adatvezérelt világban?
 - adatgazda vs. szolgáltatásgazda (a 20 éves open archive modell kiterjesztése az adatok világára)
 - “az adat az új olaj” Kérdés: nem inkább a szolgáltatás (elemzés, narratíva, értelmezés, szelekció)?
 - adatintegráció: ha sok adat van, az jó, de az elemzéshez integrációra van szükség
- adathitelesség, megbízhatóság
 - honnan van adat, mi az eredeti adatforrás, mitől lesz hiteles egy adat
 - KSH TSTAR | OSM | Google Maps | Lechner | MTA KRTK
 - milyen replikák vannak, hogyan lehet biztosítani, hogy a replika hiteles és nem korrupt legyen
 - Lechner TEIR, BME Radar
 - állami és piaci adatrendszerek jelenidejűségének problémája
 - történelmi idősorok (hiányzó évek)
 - változáskövetés, különösen megszűnéskezelés (OSM patika, vasútvonalak megszűnése – magyarázati lehetőségek elvesztése)
- adatintegráció
 - névterek használata
 - sémaszabványok
 - szemantikus tér egységesítése, ontológiák kidolgozása, propagálása (OSM mint rossz gyakorlat)
 - idő- és térbeli összemérhetőség
 - granularitás: minél kisebb területi egységre vetített, időben állandó adatkörzetek
- adattulajdonlás, adatszuverenitás
 - kié az adat?
 - nyilvánosságra kerülés = megosztás? (információs jószág)
 - el lehet-e adni közadatot (nyilvános adatot) harmadik félnek