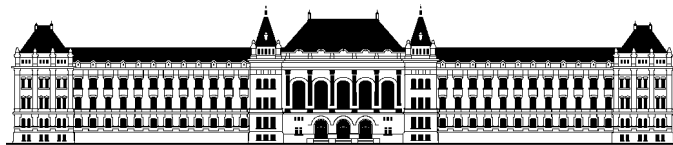


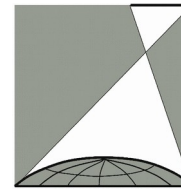
# Budapesti cégek székhelyének geokódolása az MMQGIS modul továbbfejlesztett változatával

Jankó József Attila, PhD hallgató  
Témavezető: Dr. Szabó György

Budapesti Műszaki és Gazdaságtudományi Egyetem  
Építőmérnöki Kar - Fotogrammetria és Térinformatika Tanszék



M Ű E G Y E T E M 1 7 8 2



5. Nyílt forráskódú térinformatika munkaértekezlet – BME, Budapest – 2016.11.25.

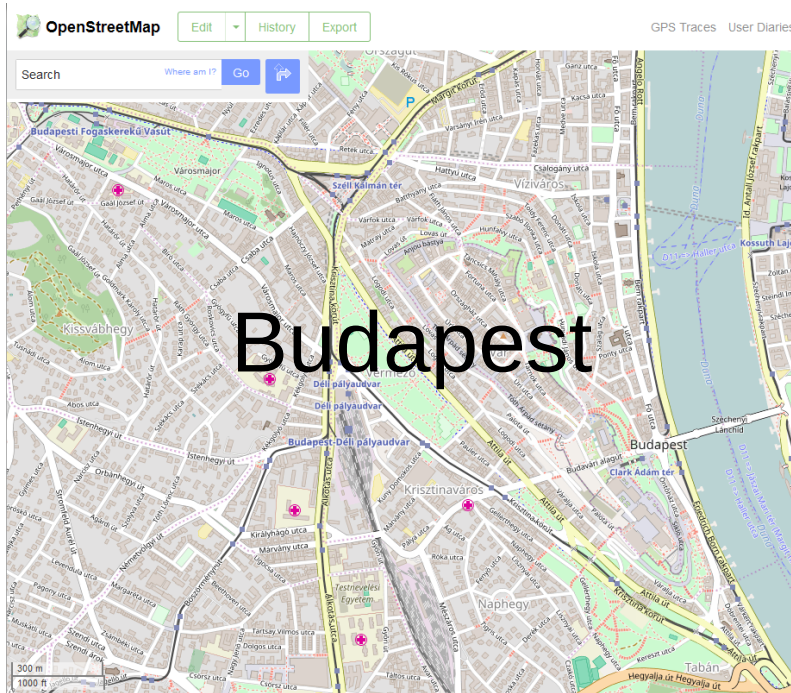
# Előzmények (2009-2010)

- Építőipari vállalkozás helyzetelőnyének élénkítése
  - (lehetséges) együttműködő partnerek **feltérképezése**
  - (tényleges) konkurencia **feltérképezése**
  - (potenciális) ügyfélkör **feltérképezése**
- Adatforrás
  - KSH (Gazdasági Szervezetek Regisztere)
    - adószám
    - statisztikai **főtevékenység** (TEÁOR'08)
    - **székhely**

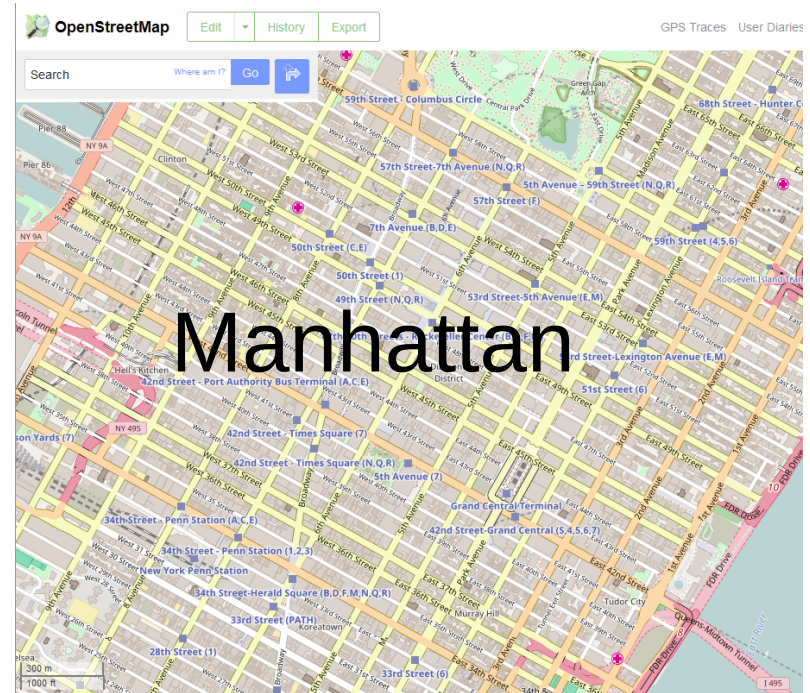
# Rendelkezésre áll ... – Tudni szeretném ...

- CSV (cím lista)
  - adószám
  - név
  - főtevékenység
  - székhely
- QGIS
  - MMQGIS
- Tudni szeretnék ...
  - valami többet
- Például ...
  - vajon kik?
  - vajon hol?
  - vajon miért?
  - vajon hogyan?

# Konfliktus



≠



# MMQGIS (Michael Minn QGIS modulja)

- Irányítószám opcionális
  - rightzipcode bug
- Házzám csak egész szám lehet
- Előzetes feltevés, hogy az utca jobb és bal oldala minden esetben páros **és** páratlan (vagy fordítva)
- A házzám-pont távolsága az utcasarkoktól egységes
- A házzám interpoláció két utcasarok közötti egyenes szakaszon történik

# MMQGIS – továbbfejlesztve

- Irányítószám opcionális, de melegen ajánlott
  - $\pm 9$  eltérés megengedhető
- Házszámokban /a /b /c /d /e /f /g lehetséges
- Vegyes paritású házszámozás (pl. tér esetén) lehetséges
- A házszám-pont távolsága az utcasarkoktól dinamikus
- A házszám interpoláció az utca-tengely mentén történik,

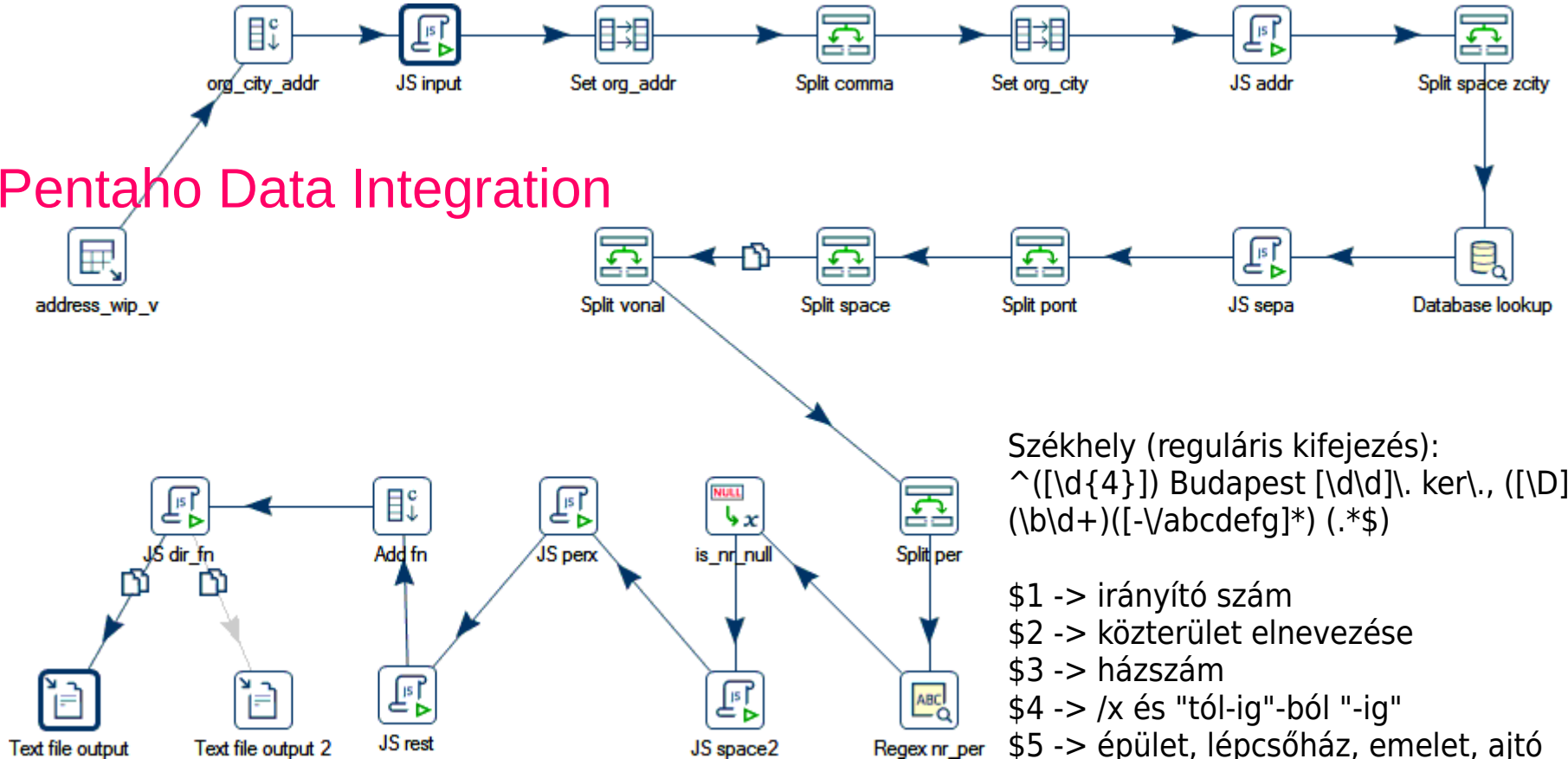
`if length_mc > length_xy * 1.001`

# MMQGIS – továbbfejlesztve<sup>2</sup>

- Ismeretlen házszámozás esetén a házszám-pont a legközelebbi utca-szakasz tengely közepére tehető
- Cím nyomdahiba (elgépelés) esetére 5 minőségi szintű közterület-név összehasonlítás választható
- A irányítószám-eltérés, a közterület-név elgépelés és a házszám beazonosítottság minőségi jelzővel regisztrálásra kerül
- Az utca-tengely objektum azonosítója regisztrálásra kerül

# CSV (cím lista) előállítás

- Pentaho Data Integration



Székhely (reguláris kifejezés):  
^([\d{4}]) Budapest [\d\d]\. ker\., ([\D]\*  
(\b\d+)([-Vabcdefg]\*) (.\*\$)

- \$1 -> irányító szám
- \$2 -> közterület elnevezése
- \$3 -> házszám
- \$4 -> /x és "tól-ig"-ból "-ig"
- \$5 -> épület, lépcsőház, emelet, ajtó



# CSV (cim lista) előállítás

- Modified Java Script Value
  - Regex Evaluation helyett is

Step name: JS input

Java script functions:

- Transform Scripts
  - Transform Constants
  - Transform Functions
  - Input fields
    - tbl\_201\_id
    - tbl\_201\_cid
    - tbl\_201\_input
    - jar
    - org\_input
    - org\_city
    - hsz
  - Output fields
    - Please use the 'Replace'

Script 1

```
tbl_201_input = tbl_201_input.replace(/(tbl_201_input)/i);
tbl_201_input = tbl_201_input.replace(/tbl_201_input/g, "tbl_201_input");

tbl_201_input = tbl_201_input.replace(/(\d)([\d\.,-])/i, "$1x8x8x$2");
tbl_201_input = tbl_201_input.replace(/x8x8x/g, "0");

tbl_201_input = tbl_201_input.replace(/(5\d\d\.)([ust])/i, "$1 $2");
tbl_201_input = tbl_201_input.replace(/(5\d\d)([sust])/i, "$1. $2");
tbl_201_input = tbl_201_input.replace(/(5\d\d)[a-zA-Z-]+([uUsStT])/, "$1. $2");

tbl_201_input = tbl_201_input.replace(/(\bu),[\s]*(\d)/i, "$1. $2");
tbl_201_input = tbl_201_input.replace(/(\bu)[\.,\s]+(\d)/i, " utca $2");

tbl_201_input = tbl_201_input.replace(/(\butca)[\.,\s]+(\d)/i, " utca $2");
tbl_201_input = tbl_201_input.replace(/(\s,u)[\s]*(\d)/i, " utca $2");
tbl_201_input = tbl_201_input.replace(/(\D)\s(u)[\s]*(\d)/i, "$1 utca $3");

tbl_201_input = tbl_201_input.replace(/(\d),(\d)/, "$1;$2");
tbl_201_input = tbl_201_input.replace(' ', ',');
tbl_201_input = tbl_201_input.replace(' ', ',');
tbl_201_input = tbl_201_input.replace(' ', ',');

tbl_201_input = tbl_201_input.replace('/', '/');
tbl_201_input = tbl_201_input.replace(/(\d)\.-(\d)/, "$1-$2");

tbl_201_input = tbl_201_input.replace(/(\w)\.(\w)/, "$1. $2");

tbl_201_input = tbl_201_input.replace('32-esek tere', 'Harminckettesek tere');
tbl_201_input = tbl_201_input.replace('32EK TERE', 'Harminckettesek tere');
tbl_201_input = tbl_201_input.replace('32-ESEK TERE', 'Harminckettesek tere');
tbl_201_input = tbl_201_input.replace('32-SEK TERE', 'Harminckettesek tere');

tbl_201_input = tbl_201_input.replace('56-osok tere', 'Ötvenhatosok tere');

tbl_201_input = tbl_201_input.replace(/Március\s15/i, 'MárciusXV');
tbl_201_input = tbl_201_input.replace(/Október\s/i, 'OktóberXI');
```

Linens: 0

Compatibility mode?  Optimization level 9

#	Fieldname	Rename to	Type	Length	Precision	Replace value 'Fieldname' or 'Rename ...
1	tbl_201_input		String			Y

# CSV (cím lista) előállítás

Step name:

- Irányítószám
- Település neve
- Közterület neve
- Házzám (első, numerikus rész, egész szám)
- *la lb lc ld le lf lg*
- Házzám (második numerikus rész)
- Épület, lépcsőház, emelet, ajtó
- Házzám (lebegőpontos)

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim Type	Null
1	tbl_201_id	Integer							both	
2	tbl_201_cid	Integer							both	
3	zip	String							both	
4	city	String							both	
5	street	String							both	
6	hsz	String							both	
7	nr_per	String							both	
8	nr_von	String							both	
9	nr_spa	String							both	
10	sub_a	String							both	
11	numero	String							both	

Get Fields Minimal width

Help

OK

Cancel

# Hátszám NEM csak egész szám lehet

```
if(nr_per == 'a'){
    numero = hsz + '.3';
} else if(nr_per == 'b'){
    numero = hsz + '.6';
} else if(nr_per == 'c'){
    numero = hsz + '.8';
} else if(nr_per == 'd'){
    numero = hsz + '.81';
} else if(nr_per == 'e'){
    numero = hsz + '.9';
} else if(nr_per == 'f'){
    numero = hsz + '.91';
} else if(nr_per == 'g'){
    numero = hsz + '.93';
} else {
    nr_per = null;
}
```

# Utca-tengely objektum (extra) mezői

- fromright = -2, toright = -2, HA ...
- fromleft/fromright = -3/-4, HA ...
- toleft/toright = -3/-4, HA ...
- Dinamikus utcasarok – házszám-pont távolság (**widthleft, widthright**) az utca-szakasz relatív házszám-pont sűrűségének függvényében

```
57 EXECUTE 'UPDATE tmp.street_cnt_line_' || kerulet || ' SET fromright = -2, toright = -2 WHERE fromright = -1 AND fromleft % 2 = 1;';
58
59 EXECUTE 'UPDATE tmp.street_cnt_line_' || kerulet || ' d SET fromleft = -3, toleft = -3 FROM (SELECT g.gml_oid as xgid
60 FROM tmp.street_cnt_line_' || kerulet || ' c, tmp.street_cnt_line_' || kerulet || ' g WHERE c.fromleft = -2 AND c.street =
g.street AND ST_Intersects(c.geom, g.geom)) s WHERE d.fromleft = -1 AND d.gml_oid = s.xgid;';
61
62 EXECUTE 'UPDATE tmp.street_cnt_line_' || kerulet || ' d SET fromright = -3, toright = -3 FROM (SELECT g.gml_oid as xgid
63 FROM tmp.street_cnt_line_' || kerulet || ' c, tmp.street_cnt_line_' || kerulet || ' g WHERE c.fromright = -2 AND c.street =
g.street AND ST_Intersects(c.geom, g.geom)) s WHERE d.fromright = -1 AND d.gml_oid = s.xgid;';
64
65 EXECUTE 'UPDATE tmp.street_cnt_line_' || kerulet || ' d SET fromleft = -4, toleft = -4 FROM (SELECT g.gml_oid as xgid
66 FROM tmp.street_cnt_line_' || kerulet || ' c, tmp.street_cnt_line_' || kerulet || ' g WHERE c.fromleft = -3 AND c.street =
g.street AND ST_Intersects(c.geom, g.geom)) s WHERE d.fromleft = -2 AND d.gml_oid = s.xgid;';
67
68 EXECUTE 'UPDATE tmp.street_cnt_line_' || kerulet || ' d SET fromright = -4, toright = -4 FROM (SELECT g.gml_oid as xgid
69 FROM tmp.street_cnt_line_' || kerulet || ' c, tmp.street_cnt_line_' || kerulet || ' g WHERE c.fromright = -3 AND c.street =
g.street AND ST_Intersects(c.geom, g.geom)) s WHERE d.fromright = -2 AND d.gml_oid = s.xgid;';
70
71 EXECUTE 'UPDATE tmp.street_cnt_line_' || kerulet || ' SET widthleft =
abs(ln(ST_Length(ST_LineMerge(geom))/10.0))/3.0*ST_Length(ST_LineMerge(geom))/GREATEST((toleft-fromleft)::float,1.0)/2.0 WHERE
toleft < -2 OR toright < -2;';
72 EXECUTE 'UPDATE tmp.street_cnt_line_' || kerulet || ' SET widthright =
abs(ln(ST_Length(ST_LineMerge(geom))/10.0))/3.0*ST_Length(ST_LineMerge(geom))/GREATEST((toright-fromright)::float,1.0)/2.0 WHERE
toright < -2 OR toleft < -2;';
73
74 EXECUTE 'UPDATE tmp.street_cnt_line_' || kerulet || ' SET widthleft =
abs(ln(ST_Length(ST_LineMerge(geom))/10.0))/3.0*ST_Length(ST_LineMerge(geom))/GREATEST((toleft-fromleft)::float,1.0) WHERE toleft
> 0 AND widthleft = 0;';
75 EXECUTE 'UPDATE tmp.street_cnt_line_' || kerulet || ' SET widthright =
abs(ln(ST_Length(ST_LineMerge(geom))/10.0))/3.0*ST_Length(ST_LineMerge(geom))/GREATEST((toright-fromright)::float,1.0) WHERE
toright > 0 AND widthright = 0;';
76
77 END IF;
78
79 i := i + 1;
80 END LOOP; -- FOREACH
```

# A továbbfejlesztett MMQGIS

Input CSV File (UTF-8)  
W:\usr\_u32\QGIS 2.18\bin\address\_list.csv

Street Name Field  
street

Street Layer  
street\_cnt\_line

Street Name Attribute  
street

Bldg. Setback (map units)  
15

Address match quality  
1

Output Shapefile  
W:\usr\_u32\QGIS2~1.18\bin\temp.shp

Not Found Output List  
W:\usr\_u32\QGIS2~1.18\bin\notfound.csv

OK Cancel

**Utcategy – házszám-pont  
távolság: 15 m vetületi  
rendszerben  
(2/3, ha  $l < 95m$ )**

**Közterület név  
találat minősége  
(aeiou = 1,2,3,4,5)**

Input CSV File (UTF-8)  
W:\usr\_u32\QGIS 2.18\bin\address\_list.csv

Street Name Field  
street

Number Field  
numero

ZIP Field  
zip

Street Layer  
street\_cnt\_line

Street Name Attribute  
street

From X Attribute  
(street line start)

From Y Attribute  
(street line start)

Bldg. Setback (map units)  
15

Address match quality  
1

Allowance for post code alterations

Allowance for uncoded house numbers

Output Shapefile  
W:\usr\_u32\QGIS2~1.18\bin\temp.shp

Not Found Output List  
W:\usr\_u32\QGIS2~1.18\bin\notfound.csv

OK Cancel

**Irányítószám  
eltérés  
megengedve**

**Házszám-pont az  
utcatengely-  
szakasz közepén**

# Geokódolás minősége

```
2255 # Build attribute fields for geocoded address shapefile
2256 fields = QgsFields()
2257 header = reader.next()
2258 header.append('qual') # jja
2259
```

Közterület név minőség	Irányítószám	Közterület név egyezés	Házzám	
<b>9,8,7,6,5,0</b>	<b>7,5,0</b>	<b>5,0</b>	<b>3,0</b>	
10-aeiou	7 - azonos 5 - ± 9	5 - egyezik	3 - hsz. interpolált	
0 - nincs találat	0 - nincs találat	0 - nincs találat	0 - nincs találat	
			0	nincs találat
		5	0	közterület létezik
9-5	7	5	3	
<b>9</b>	<b>7</b>	<b>5</b>	<b>3</b>	<b>Teljes egyezés</b>

# Utcatengely – házszám-pont távolság

15 m vetületi rendszerben  
(2/3, ha  $l < 95\text{m}$ )

```
2516  
2517  
2518  
2519  
2520  
2521  
2522   
2523  
2524  
2525   
2526
```

```
debug_indx_int[feature_indx_mc,row_indx_mc] += 3 # jja house nr matched  
row[qual_index_mc] = aeiou + debug_indx_int[feature_indx_mc,row_indx_mc]  
  
# Find line start and end points  
geometry = feature.geometry()  
length_mc = geometry.length() # jja  
if length_mc < 95:  
    # sort street might be narrower -> use smaller setback  
    setback_mc = setback / 3 * 2 # jja  
else:  
    setback_mc = setback # jja
```

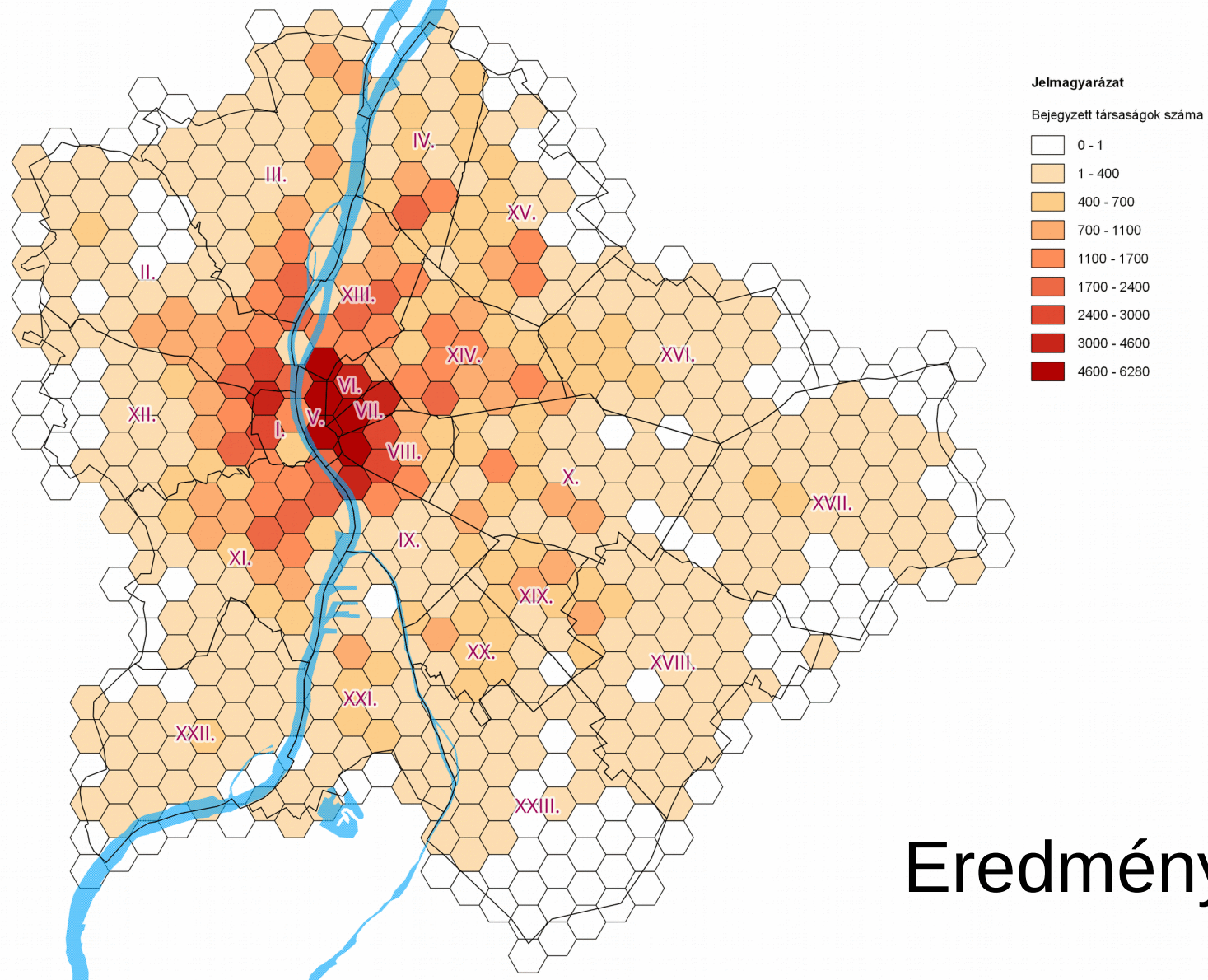
# Írányítószám eltérés megengedve

```
2473 # allowance for post code alteration
2474 if zip_allowance == 2:
2475     zip_diff = 9
2476 else:
2477     zip_diff = 0
2478 try:
2479     zipcode_int = int(zipcode)
2480     leftzipcode_int = int(leftzipcode)
2481     rightzipcode_int = int(rightzipcode)
2482     leftzip = ((zipcode_int >= leftzipcode_int - zip_diff) and (zipcode_int <= leftzipcode_int +
2483     zip_diff))
2484     rightzip = ((zipcode_int >= rightzipcode_int - zip_diff) and (zipcode_int <= rightzipcode_int +
2485     zip_diff))
2486 except:
2487     QgsMessageLog.logMessage('zipcode as was')
2488     leftzip = ((zipcode == leftzipcode))
2489     rightzip = ((zipcode == rightzipcode))
2490 if not (zipcode == rightzipcode) and not (zipcode == leftzipcode):
2491     debug_indx_int[feature_indx_mc,row_indx_mc] = 550 # jja postal code not matched exactly
2492     row[qual_index_mc] = aeiou + debug_indx_int[feature_indx_mc,row_indx_mc]
2493 if not leftzip and not rightzip:
2494     debug_indx_int[feature_indx_mc,row_indx_mc] = 50 # jja postal code not matched
2495     row[qual_index_mc] = aeiou + debug_indx_int[feature_indx_mc,row_indx_mc]
```



# Lineáris interpoláció (ha $l > AB * 1,001$ )

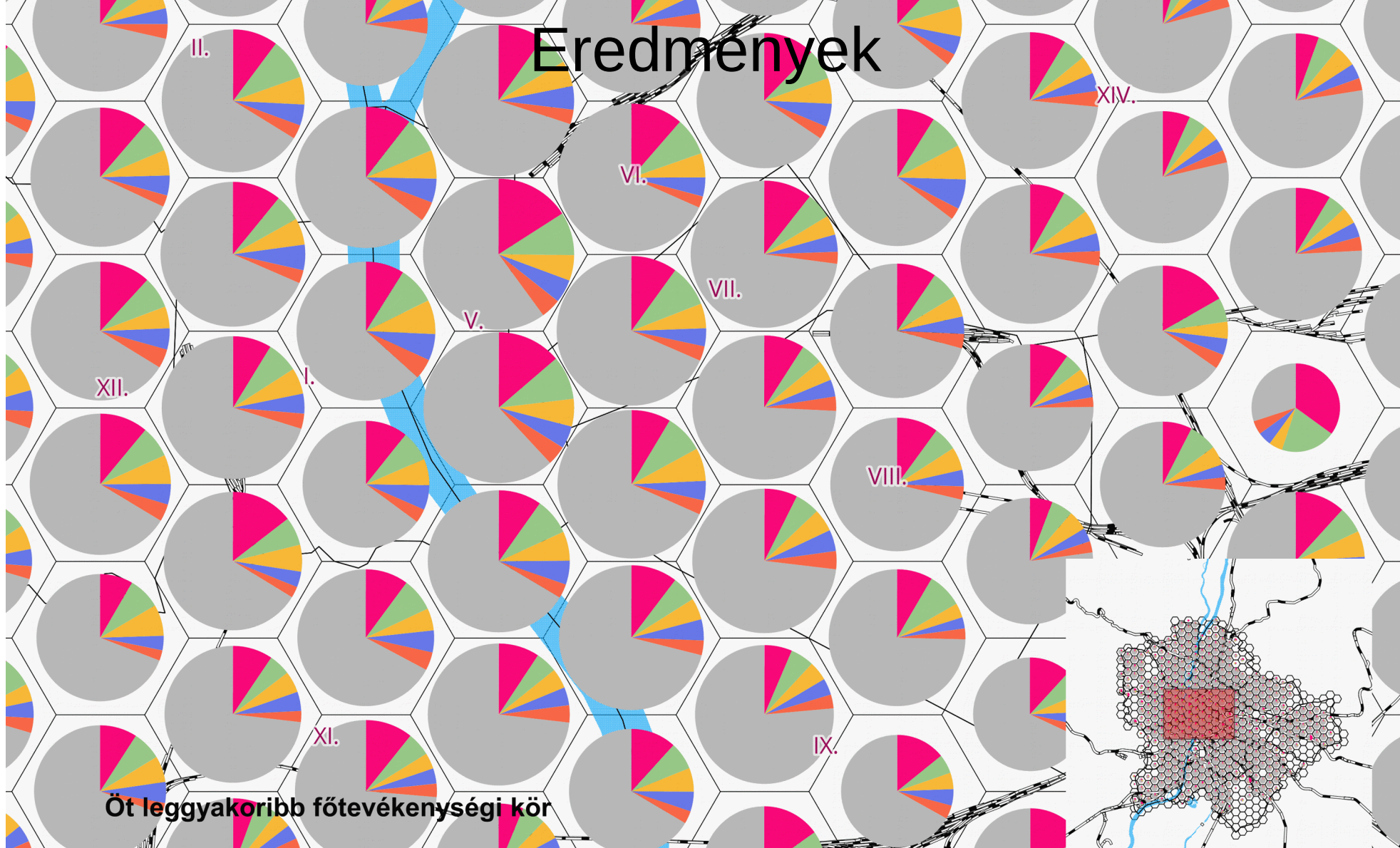
```
53 def lin_int_mc(geometry, linear_pos, oid=(0,0,0)):
54     lin_cum = 0
55     if (geometry.wkbType() == QGis.WKBLineString):
56         line = geometry.asPolyline()
57         for i in range(0, len(line)):
58             fromx = line[i].x()
59             fromy = line[i].y()
60             tox = line[i+1].x()
61             toy = line[i+1].y()
62             lin_cur = sqrt(pow(tox - fromx, 2.0) + pow(toy - fromy, 2.0))
63             if (lin_cum + lin_cur) >= linear_pos:
64                 angle = atan2(toy - fromy, tox - fromx)
65                 ratio = (linear_pos - lin_cum) / lin_cur
66                 x = fromx + ((lin_cur * ratio) * cos(angle))
67                 y = fromy + ((lin_cur * ratio) * sin(angle))
68                 return (x,y,angle)
69             else:
70                 lin_cum += lin_cur
71
72     elif (geometry.wkbType() == QGis.WKBMultiLineString):
73         lines = geometry.asMultiPolyline()
74         for m in range(0, len(lines)+1):
```



# Eredmények



# Eredmények



Öt leggyakoribb főtevékenységi kör

# Forrásjegyzék

- Ráday M. (2013): Budapesti utcanevek A-Z, Corvina Kiadó
- Internet:
  - [http://portal.ksh.hu/pls/portal/vb.teaor\\_main.gszr\\_main2](http://portal.ksh.hu/pls/portal/vb.teaor_main.gszr_main2)
  - <http://91.82.249.168:8080/geoserver/ows>
  - <http://www.qgis.org/en/site/>
  - <http://michaelminn.com/linux/mmqgis/>
  - <http://nostra.org/r/yh75d8d>

# Köszömöm a figyelmet!

- mmqgis-fork-v1 a Github-on:
  - <http://nostra.org/r/yh75d8d>