

# Testovanie priestorovej autokorelácie nezamestnanosti absolventov vysokých škôl okresov Slovenska

Dagmar Kusendová<sup>1</sup>, Jana Solčianska<sup>2</sup>

Katedra humánnej geografie a demogeografie, Prírodovedecká fakulta, Univerzita  
Komenského, Mlynská dolina,  
84215, Bratislava, Slovensko  
kusendova@fns.uniba.sk<sup>1</sup>  
jsolcianska@yahoo.com<sup>2</sup>

**Abstrakt.** Geoinformačné technológie sprístupňujú humánnym geografovi špecifické analytické metódy, ku ktorým patrí priestorová autokorelácia. Príspevok sa venuje problematike nezamestnanosti absolventov vysokých škôl na úrovni okresov Slovenska v roku 2005 riešenej metódami priestorovej autokorelácie pomocou lokálnych indikátorov priestorovej štatistiky, konkrétne lokálneho Moranovho a lokálneho G koeficientu. Pri výpočtoch a tvorbe výstupov bol použitý program Space-Time Intelligence System s implementovanými nástrojmi priestorovej štatistiky a kartografickej vizualizácie. Článok uzatvára hodnotenie a porovnanie výsledkov získaných oboma metódami.

**Kľúčové slová:** priestorová autokorelácia, lokálne indikátory, nezamestnanosť, okresy Slovenska.

**Abstract.** *Testing of the spatial autocorrelation of unemployment of high schools graduates in the districts of Slovakia.* Geoinformation technologies provide human geographers with some specific analytical methods the spatial autocorrelation belongs to. The article refers to methodical procedures of the spatial autocorrelation testing with the help of local indicators of spatial statistics applying Moran's and local G coefficients. This method was used in the case study aimed at the testing of unemployment of high schools graduates in the districts of the Slovakia in 2005. The Space-Time Intelligence System software with implemented tools of spatial statistics and cartographic visualisation was used for computing and outputting. The article evaluates and compares the results gained using both methods.

**Keywords:** spatial autocorrelation, local indicators, unemployment, districts of Slovakia

## 1 Úvod

Geoinformačné technológie sprístupňujú geografovi špecifické analytické metódy, ku ktorým patrí aj priestorová autokorelácia. Naším cieľom bolo objasniť metodiku, resp. postup testovania priestorovej autokorelácie v prostredí programu Space-Time Intelligence System [5] na vybranom humánno-geografickom probléme, a to analýze nezamestnanosti absolventov vysokých škôl v okresoch Slovenska v roku 2005.

## 2 Lokálne indikátory priestorovej štatistiky

Pojem autokorelácia označuje postup hodnotenia korelácie v rámci jednej premennej. Priestorová autokorelácia je špecifickým typom korelácie, kde v rámci jedného pozorovania sa hodnotí vzťah jednej premennej v priestore a čase. V sociálno-ekonomických javov veľmi často dochádza k priestorovej interakcii, v dôsledku čoho sa môžu podobné javy alebo atribúty nachádzať v priestore bližšie (pozitívna autokorelácia), alebo dochádza k zoskupeniu výrazne odlišných hodnôt (negatívna autokorelácia). Z geografického hľadiska sa priestorová autokorelácia hodnotí ako vzťah medzi javmi alebo udalosťami oddelenými určitými priestorovými alebo časovými úsekmi. Podľa Stehlíkovej [4] je model priestorovej autokorelácie priestorovým ekvivalentom modelu časového radu, rozdiel je len v tom, že závislosť v čase je spätná, zatiaľ čo závislosť v priestore je viacsmerová. Na hodnotenie priestorovej autokorelácie sa vypracovalo množstvo nástrojov. Hlásny [2] tieto nástroje člení na globálne indikátory priestorovej asociácie, lokálne indikátory a geoštatistické nástroje založené na báze variogramov.

V analýze sme sa sústredili len na tzv. *lokálne indikátory priestorovej štatistiky*, ktoré sa dajú použiť na identifikovanie oblastí, kde susedné územia majú podobné hodnoty alebo území, ktoré sú odlišné od svojich susedov. Konkrétne sme použili *lokálny Moranov koeficient* a *G štatistiku*, ktoré patria k všeobecne najpoužívanejším, pričom sme sa zamerali na hodnotenie a porovnanie výsledkov.

Pri výpočtoch a tvorbe výstupov použitý program *Space-Time Intelligence System* (ďalej STIS) disponuje vybranými metódami priestorovej štatistiky, ktoré sú orientované predovšetkým na aplikácie z oblasti sociálno-ekonomických vied. Keďže STIS poskytuje dobrú kartografickú vizualizáciu výsledkov dá sa uplatniť aj v humánnej geografii, resp. demogeografii.

### 2.1 Lokálna Moranova štatistika

Moranov a Gearyho koeficient, ako lokálne indikátory priestorovej asociácie, zaviedol v roku 1995 Anselin [4]. Moranov lokálny koeficient sa dá aplikovať na pravidelnú alebo nepravidelnú sieť polí, resp. priestorových jednotiek výskytu javu, čo zahŕňa aj administratívne jednotky, pričom môžu byť použité nominálne, ordinálne alebo intervalové dáta [3] vo vzťahu:

$$I_i(d) = \frac{x_i - \bar{x}}{\frac{\sum_{j=1; j \neq i}^n w_{ij}(d)}{n-1} - \bar{x}^2} \sum_{j=1; j \neq i}^n w_{ij}(d)(x_j - \bar{x})$$

$d$  – kritická vzdialenosť,

$n$  – počet priestorových jednotiek,

$x_i$  ( $i=1,2,\dots,n$ ) – hodnota skúmaného javu v priestorovej jednotke  $i$ ,

$\bar{x}$  – priemerná hodnota skúmaného znaku,

$w_{ij}(d)$  ( $i, j = 1, 2, \dots, n$ ) – váha pre jednotku  $i, j$  a vzdialenosť  $d$ .

STIS implementuje dva typy lokálnej Moranovej štatistiky: jedno- a dvojvariantný. Prvý slúži na odhaľovanie priestorovej autokorelácie jednej premennej v jednom období, druhý testuje priestorovú autokoreláciu dvoch premenných v jednej časovej perióde alebo priestorovo-časovú autokoreláciu jednej premennej v čase.

Vo výpočte Moranovho koeficientu program využíva metódu *Monte Carlo* na tzv. podmienenú „randomizáciu“, ktorá prerozdeľuje chybnú početnosť hodnôt medzi priestorovými jednotkami. V každej randomizácii je chybná početnosť viazaná na jednu priestorovú jednotku a zvyšné hodnoty sú novo náhodne lokalizované. Náhodnosť je teda podmienená, pretože všetkým jednotkám s výnimkou jednej sú priradené náhodné početnosti. Tento proces sa potom opakuje pre každú z nich.

Podstatou použitej metódy je testovanie nulovej hypotézy  $H_0$ , ktorá znie: medzi hodnotami pozorovanými v okrese  $i$  a hodnotami pozorovanými v blízkyoch okresoch nie je žiadna asociácia.

Alternatívna hypotéza  $H_a$  predpokladá, že hodnoty v blízkyoch okresoch sú podobné alebo rozdielne ako hodnota v okrese  $i$ .

## 2.2 Moranov diagram rozptylu

Moranov diagram rozptylu (obr. 1) poskytuje názornú ukážku priestorovej autokorelácie. Diagram je automaticky generovaný programom pri výpočte lokálneho Moranovho koeficientu. Anselin [1] diagram opisuje ako „priestorové oneskorenie premennej na vertikálnej osi a pôvodnej premennej na horizontálnej osi“. Priestorové oneskorenie predstavujú hodnoty susedov tej ktorej priestorovej jednotky.

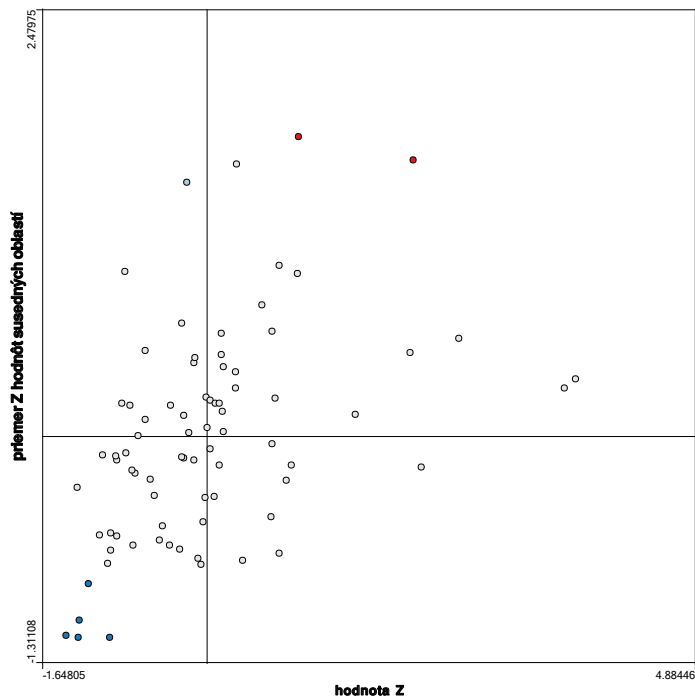
Diagram rozptylu zobrazuje normalizované dáta, ktoré predstavujú podiel rozdielu medzi pozorovanými a očakávanými (t.j. strednými) hodnotami premennej k štandardnej odchýlke na základe výpočtového vzťahu:

$$Z = \frac{x_{i,t} - \bar{x}_t}{s_{v,t}}$$

$x_{i,t}$  – premenná v priestorovej jednotke  $i$  a čase  $t$ ,

$\bar{x}_t$  – priemer premenných  $x$  v čase  $t$  vo všetkých priestorových jednotkách,

$s_{v,t}$  – štandardná odchýlka premennej  $x$  v čase  $t$ .



Obr. 1. Diagram rozptylu<sup>2</sup>

Body prislúchajúce jednotlivým priestorovým jednotkám nadobúdajú v diagrame rôzne farby: sivé sú štatisticky nevýznamné, červené a tmavomodré sú pozitívne autokorelované, ružové a slabomodré negatívne autokorelované.

### 2.3 Lokálna G štatistika (Getisova)

Lokálne  $G_i(d)$  a  $G_i^*(d)$  štatistiky opísali v roku 1995 Getis a Ord [4]:

$$G_i(d) = \frac{\sum_{j=1; j \neq i}^n w_{ij}(d)x_j - \bar{x}_i \sum_{j=1; j \neq i}^n w_{ij}(d)}{S_i \sqrt{\frac{(n-1) \sum_{j=1; j \neq i}^n w_{ij}^2(d) - \left( \sum_{j=1; j \neq i}^n w_{ij}(d) \right)^2}{N-2}}}$$

<sup>2</sup> Vysvetlivky k diagramu:

*x*-ová os – normalizované hodnoty pre každú priestorovú jednotku,

*y*-ová os – priemer štandardizovaných hodnôt susedov,

*počiatok* – zobrazené dáta sú normalizované, takže ich stredná hodnota je nula a osi sa

pretínajú v bode nula,

*sklon* – sklon úsečky, ktorá by sa dala preložiť zobrazenými dátami, je úmerný globálnemu

Moranovmu koeficientu skúmaných dát.

kde

$$\bar{x}_i = \frac{\sum_{j=1; j \neq i}^n x_j}{n-1} \quad S_i = \sqrt{\frac{\sum_{j=1; j \neq i}^n x_j^2}{n-1} - (\bar{x}_i)^2}$$

$d$  – kritická vzdialenosť,

$n$  – počet priestorových jednotiek,

$x_i$  ( $i=1,2,\dots,n$ ) – hodnota skúmaného javu v priestorovej jednotke  $i$ ,

$w_{ij}(d)$  ( $i, j = 1, 2, \dots, n$ ) – váha pre jednotku  $i, j$  a vzdialenosť  $d$ .

Obe štatistiky indikujú v akom rozsahu je lokalita  $i$  obklopená zhlukom vysokých, resp. nízkych hodnôt skúmaného znaku. Na rozdiel od lokálnej Moranovej štatistiky, hodnotí lokálna  $G$  štatistika len pozitívnu priestorovú autokoreláciu.

Nulová hypotéza  $H_0$  na testovanie štatistiky znie: okolo okresu  $i$  nie je žiaden zhluk vysokých alebo nízkych hodnôt, testovacia štatistika je blízka nule; zatiaľ čo alternatívna hypotéza  $H_a$  znie: okolo okresu  $i$  je zhluk vysokých alebo nízkych hodnôt, ktorého je okres  $i$  súčasťou. Štatisticky významné pozitívne hodnoty naznačujú zhluk vysokých hodnôt a štatisticky významné negatívne hodnoty indikujú zhluk nízkych hodnôt.

Štatistická významnosť každej  $G_i$  hodnoty je odhadovaná pomocou metódy *Monte Carlo*, ktorá simuluje pravdepodobnosť výskytov stavov u jednotlivých variant a používa vysoký počet opakovaní na dosiahnutie optimálnych výsledkov.

Vo všeobecnosti sa metóda skladá z týchto krokov [5]:

- po výpočte štatistiky z pôvodných dát sú dáta náhodne preusporiadané,
- štatistika je prepočítavaná pre náhodne preusporiadané dáta,
- prvé dva kroky sú opakované stanovený počet krát, pričom sa zhromažďujú rozdelenia, ktoré sú ďalej použité na výpočet  $P$ -hodnôt pre skúmané dáta,
- $P$ -hodnoty sú počítané porovnávaním štatistiky z pôvodných dát k príslušnému rozdeleniu.

## 2.4 Priestorové váhy

Pri odhadovaní priestorovej asociácie v dátach sa najprv musia definovať priestorové vzťahy medzi územnými jednotkami/oblasťami. Opisujú ich priestorové váhy počítané z bodových alebo polygónových dát, pričom sa odlišuje spôsob výpočtu. Prvým krokom pri analyzovaní dát je generovanie matice susedstva územia (obr. 2).



$$W = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

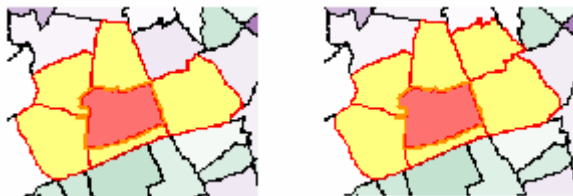
**Obr. 2.** Priestorové rozloženie polygónov v území a generovaná matica susedstva

Ak sa susedstvo nemeria, všeobecne sa dajú priestorové váhy definovať nasledovne:

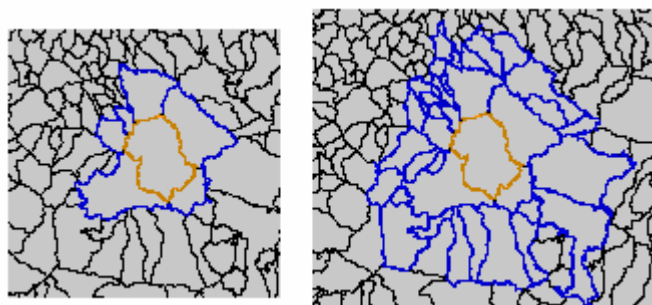
$W_{ij} = 1$ , ak oblasť  $j$  susedí s oblasťou  $i$

$W_{ij} = 0$ , ak oblasť  $j$  nesusedí s oblasťou  $i$  alebo  $i = j$  [4].

V použitom programovom prostredí sú pre polygónové dáta voliteľné rôzne typy váh, ktoré sú vzájomne kombinovateľné (obr.3–5).

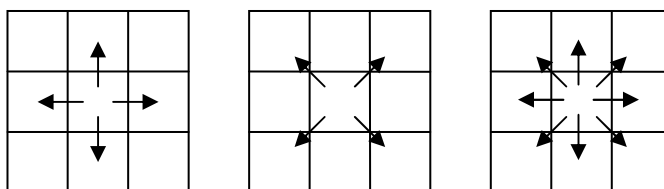


Obr. 3. Priestorové váhy typu Rook a Queen.



Obr. 4. Priestorové váhy 1. a 2. stupňa.

Váhy typu *Rook* a *Queen* predstavujú priestorové štruktúry. V prvom prípade sa do úvahy berú len oblasti susediace s oblasťou  $i$  na hranách, zatiaľ čo pri váhach typu *Queen* sa počíta so všetkými susednými oblasťami. Okrem toho existujú aj váhy typu *Bishop*, u ktorých sa uvažuje len s oblasťami susediacimi na vrcholoch.



Obr. 5. Zovšeobecnená priestorová štruktúra váh typu: Rook, Bishop a Queen.

### 3 Použité premenné

V testovaní priestorovej autokorelácie nezamestnanosti absolventov vysokých škôl v okresoch Slovenska boli použité dáta z pravidelných štatistických zisťovaní. Jediným dostupným zdrojom štatistických údajov o absolventskej nezamestnanosti sú údaje z Ústredia práce, sociálnych vecí a rodiny [6], ktoré mapujú problematiku tzv. evidovanej nezamestnanosti. Tá zachycuje počet tých, ktorí sa po strate zamestnania

prihlásia a zaevidujú na príslušnom úrade práce. Problémom štatistiky Ústredia práce, sociálnych vecí a rodiny je, že mimo nej zostáva takmer 20 % nezamestnaných, ktorí si zamestnanie hľadajú bez pomoci úradov práce.

Štatistiky o nezamestnanosti absolventov sú spracovávané dvakrát do roka, k 31. 5. a k 30. 9 v územnej mierke Slovenskej republiky, krajov a okresov.

Analýza sa zameriava na testovanie priestorovej štatistickej závislosti počtov nezamestnaných absolventov vysokých škôl k 30.9 na úrovni okresov v roku 2005. Použitým analytickým ukazovateľom bol podiel počtu uchádzačov o zamestnanie, ktorí sú absolventmi vysokých škôl k danému termínu z počtu disponibilných uchádzačov o zamestnanie v %. Uchádzač o zamestnanie je v našom prípade nezamestnaný absolvent vysokej školy mladší ako 25 rokov, ktorý skončil sústavnú prípravu na povolanie v dennej forme štúdia pred menej ako dvomi rokmi a nezískal prvé pravidelne platené zamestnanie. Pod disponibilným uchádzačom o zamestnanie sa rozumie uchádzač, ktorý bezprostredne po ponuke voľného pracovného miesta môže nastúpiť do pracovného pomeru.

V prípade oboch koeficientov sme testovali nulovú hypotézu, že v okresoch Slovenska nie je žiadna priestorová závislosť medzi pomermi absolventov vysokých škôl k disponibilným nezamestnaným.

## 4 Závěry

V skúmaní závislosti priestorového výskytu nezamestnaných absolventov vysokých škôl a nezamestnanosti celkovo, pomocou *lokálnej Moranovej* (obr. 6 – horná mapa) a *G štatistiky* (obr. 6 – dolná mapa) na úrovni okresov Slovenska, sa v niektorých regiónoch odhalila pozitívna, ale aj negatívna priestorová autokorelácia.

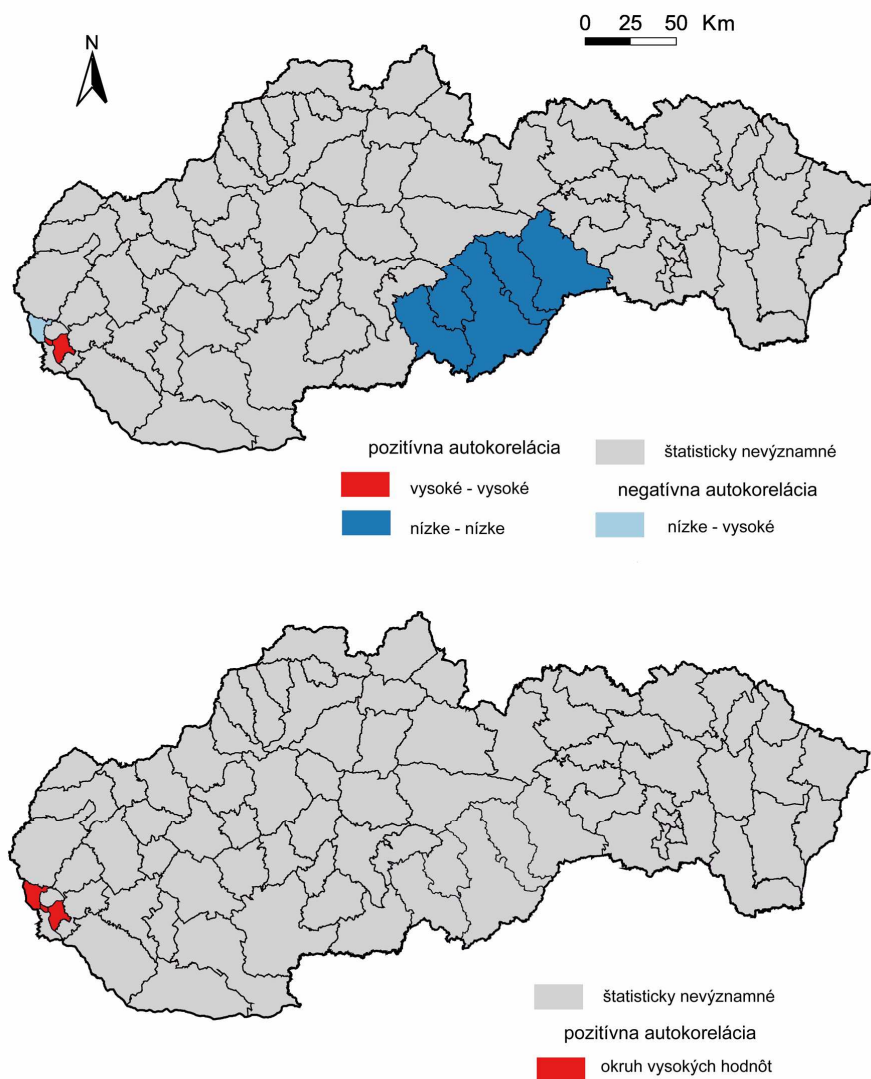
Na hladine významnosti 0,5 % a pri použití priestorových váh 1.stupňa Queen štandardizovaných podľa počtu susedov sa v prípade *lokálnej Moranovej štatistiky* vytvorili dva zhľuky pozitívnej autokorelácie. Prvý zhľuk tvorili mestské okresy Bratislava I a II. Druhý zhľuk pozitívnej autokorelácie iného typu tvorili okresy Lučenec, Poltár, Rimavská Sobota, Revúca a Rožňava.

Lokálny Moranov koeficient „odhalil“ dva známe regióny, v ktorých sa prejavujú disparity nezamestnanosti. Na jednej strane je to región Bratislavy s vysokou zamestnanosťou a na strane druhej región južného Slovenska s vysokou nezamestnanosťou. Negatívna autokorelácia bola zaznamenaná v mestskom okrese Bratislava IV. Uvedený okres sa vyznačoval nízkymi počtami nezamestnaných absolventov k disponibilným nezamestnaným, a naopak ich vysokými hodnotami v blízkom okolí.

Pri testovaní priestorovej autokorelácie pomocou *G štatistiky* sme na tej istej hladine významnosti (0,5 %) s použitím tých istých váh (1.stupňa Queen štandardizované podľa počtu susedov) dospeli k vysokým hodnotám pozitívnej autokorelácie v bratislavských mestských okresoch I, II. Pozitívne hodnoty *G štatistiky* dosahoval aj okres Bratislava IV.

Lokálna Moranova štatistika a lokálna *G štatistika* vykazujú na prvý pohľad pomerne odlišné výsledky. Rozdiely v prípade bratislavských okresov, ako aj štatistická

významnosť, resp. nevýznamnosť okresov južného Slovenska, zrejme vyplývajú z hraničnej polohy spomínaných okresov.



**Obr. 6.** Mapová vizualizácia výsledkov testovania priestorovej autokorelácie

## Reference

1. Anselin, L. *Exploring Spatial Data with DynESDA2*. CSISS and Spatial Analysis Laboratory University of Illinois, 2002, Urbana-Champaign..



2. Hlásny, T. Geoštatistický koncept priestorovej závislosti pre geografické aplikácie. *Geografický časopis*, 57. 2005.
3. Robinson, G., M. *Methods and Techniques in Human Geography*. Wiley & Sons, 1998, New York.
4. Stehlíková, B. *Priestorová štatistika*. Slovenská poľnohospodárska univerzita, 2002, Nitra.
5. *Space-Time Intelligence System (Terraseer)*. [www.terraseer.com](http://www.terraseer.com).
6. *Štatistické výsledky o nezamestnanosti absolventov škôl a mladistvých k 30. 9. 2005* (Ústredie práce, sociálnych vecí a rodiny). [www.upsvar.sk](http://www.upsvar.sk).