

KATALÓG POŽIADAVIEK

WORLD VALUE SURVEY (DEKK)

Získanie údajov, vizualizácia, výpočet korelácie,
aktualizácia dát, štatistické údaje

Obsah

1. Úvod	3
1.1 Účel katalógu požiadaviek	3
1.2 Rozsah využitia systému	3
1.3 Slovník pojmov	3
1.4 Odkazy a referencie	4
1.5 Prehľad nasledujúcich kapitol	4
2. Všeobecný popis.....	4
2.1 Perspektíva projektu	4
2.2 Funkcie systému	4
2.3 Charakteristika používateľov	4
2.4 Predpoklady, závislosti a obmedzenia	5
3. Špecifikácia požiadaviek	5
3.1 Funkčné požiadavky	5
3.1.1 Výber a prehľad datasetov	5
3.1.3 Zobrazenie grafov	6
3.1.4 Import údajov	6
3.2 Kvalitatívne požiadavky	7
3.2.1 Dlhodobosť	7
3.2.2 Jednoduchosť	7
3.2.3 Rýchlosť	7
3.2.4 Modulárnosť	7
3.2.5 Stabilita a spoľahlivosť	7
3.3 Požiadavky grafického rozhrania a mapy	7
3.4 Požiadavky používateľského rozhrania	8
3.6 Požiadavky databázového managementu	8
4. Prílohy.....	8
4.1 Zoznam používaných datasetov	9

1. Úvod

1.1 Účel katalógu požiadaviek

Dokument slúži na špecifikáciu všetkých požiadaviek k informačnému systému vytváraného pre World Value Survey (DEKK), ktorý vzniká v rámci predmetu Tvorba Informačných Systémov (1-AIN-131) na FMFI UK. Katalóg požiadaviek je určený pre všetkých stakeholderov a taktiež je záväzným pre zadávateľov projektu a jeho vývojárov. Katalóg slúži ako východiskový materiál pre následnú implementáciu projektu. Je spísaný zrozumiteľne, jednoduchým jazykom pre obe strany, tak aby došlo k minimalizácii nedorozumení medzi zadávateľmi a vývojármi.

1.2 Rozsah využitia systému

Jedná sa o projekt v oblasti geo data science s reálnym využitím v rámci výskumu sociálnej kohézie. Projekt sa snaží prispieť k riešeniu problému nevhodnej sociálnej kohézie medzi ľuďmi, ktorý sa postupne zhoršuje a ktorý je pomerne ťažko analyzovateľný. Sociálna kohézia súvisí s kriminalitou, s platmi a s veľmi veľa inými faktormi. Hlavným cieľom je začať na Slovensku, no projekt by mal byť rozšíriteľný aj do sveta.

1.3 Slovník pojmov

- **World Value Survey** - globálny výskumný projekt, ktorý skúma hodnoty a viery ľudí, ako sa v priebehu času menia a aký majú sociálny a politický dopad.
- **DEKK Inštitút** - reprezentant World Value Survey na Slovensku
- **Stakeholder** - zainteresovaná strana/“zúčastnený“
- **Data science** - interdisciplinárny obor, ktorý využíva vedecké metódy, procesy, algoritmy a systémy pre získavanie znalostí a poznatkov z dát v rôznych podobách, ako štruktúrovaných, tak neštruktúrovaných
- **Dataset** - jeden dátový atribút podľa okresov za jeden konkrétny rok, môže však byť parametrizovateľný niekoľkými atribútmi, napr. počet všetkých obyvateľov je základný atribút a je parametrizovateľný vekom a pohlavím (napr. vek=10, pohlavie=M)
- **Databanka** - úložisko informácií o jednom alebo viacerých subjektoch - databáze - ktorá je organizovaná spôsobom, ktorý uľahčuje miestne alebo vzdialené získavanie informácií a je schopný dlhodobo spracovávať mnoho neustálych dotazov
- **Geo data science** - disciplína data science zameraná špecificky na jej priestorový komponent
- **Sociálna kohézia** - charakteristika sociálnych pút a vzájomných väzieb členov spoločenskej skupiny alebo organizácie, daná prevahou spoločných záujmov/súdržnosť
- **Policy-maker** - skupina osôb/osoba zodpovedná za vytváranie a opravu nápadov/plánov/politiky
- **Slider** - posuvník/lišta/grafický ovládací prvok

1.4 Odkazy a referencie

- Odkaz na github repozitár: <https://github.com/TIS2021-FMFI/dekk>
- DEKK Inštitút: <https://dekk.sk/sk/>
- World Value Survey: <https://www.worldvaluessurvey.org/wvs.jsp>
- DATACUBE <http://datacube.statistics.sk>

1.5 Prehľad nasledujúcich kapitol

V nasledujúcej kapitole je všeobecne popísaný projekt / informačný systém - jeho perspektíva, funkcia, funkcionalita, charakteristika, predpoklady, závislosti a obmedzenia používania a v tretej kapitole je podrobná špecifikácia kvalitatívnych, funkčných požiadaviek. Taktiež pojednáva o grafickom rozhraní. Štvrtá kapitola popisuje aké dokumenty budú poskytnuté k aplikácii a obsahuje taktiež zoznam odkazov na použité datasety.

2. Všeobecný popis

2.1 Perspektíva projektu

Projekt má ambíciu byť pomocným nástrojom na analýzu zhoršujúcej sa sociálnej kohézie na Slovensku. Má slúžiť ľuďom, ktorí majú moc rozhodovať, teda politikom, novinárom, policymaker-om, ktorým pomôže pomocou vizualizácie rôznych datasetov analyzovať príčiny zlej sociálnej kohézie pre možné následné hľadanie riešenia tohto problému. Všetky datasety, s ktorými aplikácia pracuje, obsahujú jeden údaj za každý okres a určitý rok. Aplikácia sa bude vedieť vysporiadať aj s chýbajúcimi údajmi (napr. za niektorý okres v niektorom roku).

2.2 Funkcie systému

Výsledným produktom je webová aplikácia, prístupná cez web DEKK Inštitútu, ktorá umožňuje svojmu užívateľovi prehľadne graficky vizualizovať korelácie rôznych datasetov na mape Slovenska rozdelenej na okresy. Datasety bude získavať hlavne z verejných Slovenských databánk.

K funkciám aplikácie patrí aj filtrovanie v ponuke datasetov, vizualizácia zvolených datasetov na mapke Slovenska a ich prekrývania v oblastiach mapky, automatická aktualizácia dát každý rok a vykreslenie grafu na bočnom paneli pre dané filtre.

2.3 Charakteristika používateľov

Bežný používateľ - môže si zobrazovať mapku, filtrovať, prehliadať všetky dáta na všetkých paneloch. Príkladom bežného používateľa sú zamestnanci, analytici štátnej správy, akademici, študenti a časom aj novinári a široká verejnosť. Tí sa na našu databázu a jej vizuálny výstup môžu obrátiť pri ich práci.

Administrátor - má navyše možnosť na strane servera manuálne nasadiť dáta do aplikácie.

2.4 Predpoklady, závislosti a obmedzenia

Medzi obmedzenia používania aplikácie patrí: prístup na internet, funkčnosť stránky DEKK Inštitútu, cez ktorú k nej bude umožnený prístup a funkčnosť a prístup k dátovému centru DATACUBE, taktiež predpokladá, že dáta budú do aplikácie poskytované v .csv formáte.

3. Špecifikácia požiadaviek

Požiadavky označené * sú požiadavky s nižšou prioritou a budú implementované, iba ak vývojový tím bude mať kapacitu.

3.1 Funkčné požiadavky

3.1.1 Výber a prehľad datasetov

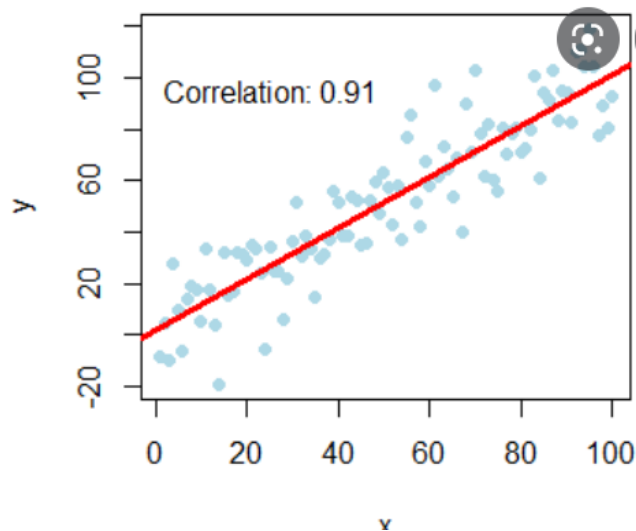
- a) V aplikácii sa bude zobrazovať zoznam prístupných datasetov, každý je určený svojím názvom
- b) Používateľ môže zoznam datasetov filtrovať – zúžiť zoznam zobrazených na také, ktorých názov obsahuje nejaký (užívateľom zadaný) textový podreťazec
- c) V zozname datasetov môže vybrať po jednom (aj pri rôznom zobrazení filtra) dva konkrétne datasety na zobrazenie korelácie, ktoré budú pre užívateľa zaujímavé
- d) * Ak dataset je parametrizovateľný, po jeho zvolení sa zobrazí pre každý parameter: názov parametra a vedľa neho drop-box v ktorom sa dá označiť jeden alebo viac riadkov. Takéto upresnenie datasetu sa použije pre zobrazenie a výpočet korelácie. (napríklad rozdelenie na mužov a ženy)

e) 3.1.2 Zobrazenie mapy

- a) Vedľa zoznamu sa zobrazuje mapa Slovenska s vyznačenými hranicami okresov, pokiaľ nie sú prístupné/označené žiadne datasety tak mapka bude biela.
- b) Po výbere dvoch datasetov používateľom sa mapa Slovenska zafarbí podľa daných hodnôt (pre každý dataset jednou farbou a jej odtieňmi).
- c) Pri mape bude zobrazená legenda, ktorá bude znázorňovať príslušnosť farieb k datasetom.
- d) Pod mapou bude zobrazený slider, ktorý umožní zobrazíť dáta z vybraného roku.
- e) Pri mape je tlačidlo umožňujúce stiahnuť obrázok aktuálne zobrazenej mapy.

3.1.3 Zobrazenie grafov

- a) Aplikácia bude zobrazovať graf, tvorený z dvoch datasetov (na x-ovej osi prvý, na y-ovej osi druhý dataset)
- b) Predbežná podoba grafu:
Každý zobrazený bod grafu zodpovedá jednému okresu.
Súčasťou zobrazenia bude celkový korelačný koeficient pre celé dáta vybranej dvojice datasetov.



- c) * Výpočet korelácie iného typu ako lineárnej.

3.1.4 Import údajov

- a) Aplikácia bude poskytovať rozhranie pre administrátora (napr. samostatnú aplikáciu spúšťanú z príkazového riadka), ktorým je možné pridať nový dataset do databázy aplikácie. Používateľ zadá meno .csv súboru, názov datasetu a rok. Ak nezadá rok, predpokladá sa aktuálny rok.
- b) Aplikácia bude poskytovať rozhranie pre administrátora (napr. samostatnú aplikáciu spúšťanú z príkazového riadka), ktorým je možné pridať nový dataset do databázy aplikácie automaticky downloadnutý zo servera Štatistického úradu SR. Používateľ zadá kód datasetu a rok. Ak nezadá rok, predpokladá sa predchádzajúci rok (napr. ak sa program pustí 1. januára 2022, tak sa downloadnú údaje z roku 2021). Typické použitie tohto programu je z automatických skriptov, ktoré ho pustia napríklad jeden raz ročne pre všetky požadované datasety.
- c) * Import popísaný 3.1.4 b. importuje kompletný dataset pre všetky hodnoty voliteľných parametrov.

3.2 Kvalitatívne požiadavky

3.2.1 Dlhodobosť

Chceme používať perspektívne platformy a nástroje tak, aby na nich bolo možné dlhodobo budovať. Nechceme meniť celú architektúru každých pár rokov.

3.2.2 Jednoduchosť

Programátorské praktiky, ktoré budú udržateľné a jednoducho pochopiteľné pre všetkých (študentské projekty stavajúce na našom systéme, ďalší administrátor, nový developer).

3.2.3 Rýchlosť

Je dôležité, aby sa mapky pre užívateľa nenačítali dlho a aby bola stránka rýchlo dostupná.

3.2.4 Modulárnosť

Malo by byť možné jednoducho pridávať ďalšie nástroje. Popríklad staré vymeniť za nové bez zbytočne komplexných závislostí medzi modulmi. To platí zvlášť pre možnosť pridávať filtre.

3.2.5 Stabilita a spoľahlivosť

Systém by nemal padať pri načítaní mapových filtrov a datasetov.

3.3 Požiadavky grafického rozhrania a mapy

Aplikácia je prehľadná, grafy, štatistické údaje a mapa sú zrozumiteľné a rozdelené do jednotlivých panelov, korelácie na nej sú zobrazované a vysvetlené jednoznačne a jednoducho. Zvolené/aplikované datasety sú farebne rozlíšené od tých nezvolených. Grafy s vyobrazenými vrstvami dát korešpondujú.

3.4 Požiadavky používateľského rozhrania

Používateľské rozhranie je intuitívne, prehľadné a užívateľovi sú jednoducho a rýchlo prístupné informácie o daných vizuálnych častiach aplikácie a ich fungovaní, taktiež je užívateľovi prístupný stručný návod na používanie priamo v aplikácii.

3.6 Požiadavky databázového managementu

Aplikácia má byť schopná uchovávať a pridávať veľké množstvo dát z viacerých rôznych databáz/datasetov. Navyše by mala periodicky-ročne preberať dáta zo zdrojov, ktoré sú pravidelne aktualizované. Tie budú pochádzať z vlastnej výskumnej produkcie (bude ich možné importovať vo formáte .csv), ale napríklad aj z dát od ministerstiev, štatistického úradu, DATACUBE.

4. Prílohy

K aplikácii bude dodaná dokumentácia aj používateľský manuál a prezentácia zamýšľaného využitia/postupu užívania aplikácie.

4.1 Zoznam používaných datasetov

Filtre z DataCube Štatistického úradu (na tieto treba pipeline)

- základná geografická úroveň - okres
- základný časový rad - ročne
- filtre:

1. Počet obyvateľov podľa pohlavia - SR-oblasť-kraj-okres, m-v (ročne) [om7102rr]
http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7102rr/v_om7102rr_00_00_00_sk

2. Živonarodení podľa poradia - SR-oblasť-kraj-okres, m-v [om7029rr]
http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7029rr/v_om7029rr_00_00_00_sk

3. Priemerný vek osoby pri úmrtí podľa pohlavia - SR-oblasť-kraj-okres, m-v [om7038rr]
http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7038rr/v_om7038rr_00_00_00_sk

a

Zomrelí podľa veku a pohlavia - SR-oblasť-kraj-okres, m-v [om7034rr]
http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7034rr/v_om7034rr_00_00_00_sk

4. Vekové zloženie - SR-oblasť-kraj-okres, m-v [om7009rr] -
http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7009rr/v_om7009rr_00_00_00_sk

5. Prehľad stavu a pohybu obyvateľstva - SR-oblasť-kraj-okres, m-v [om7011rr]
http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7011rr/v_om7011rr_00_00_00_sk

6. Veľkostné skupiny obcí
http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7023rr/v_om7023rr_00_00_00_sk

7. Stredná dĺžka života
http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7023rr/v_om7023rr_00_00_00_sk

8. Hustota obyvateľstva - SR-oblasť-kraj-okres, m-v [om7015rr]
http://datacube.statistics.sk/#!/view/sk/VBD_DEM/om7015rr/v_om7015rr_00_00_00_sk

9. Vzdelanie - počty absolventov stredných a vysokých škôl podľa okresu
http://datacube.statistics.sk/#!/view/sk/VBD_SK_WIN/sv3012rr/v_sv3012rr_00_00_00_sk
http://datacube.statistics.sk/#!/view/sk/VBD_SK_WIN/sv3024rr/v_sv3024rr_00_00_00_sk

10. Nezamestnanosť podľa rokov podľa okresov
https://www.upsvr.gov.sk/statistiky/nezamestnanost-mesacne-statistiky.html?page_id=1254
11. Priemerné platy podľa rokov podľa okresov
https://slovak.statistics.sk/wps/wcm/connect/910e7175-54da-4742-845e-19faee5b0314/Zamestnanci_a_priemerne_mesacne_mzdy_2018.pdf?MOD=AJPERES&CACHEID=ROOTWORKSPACE-910e7175-54da-4742-845e-19faee5b0314-mCIEyAL
12. Pracovná neschopnosť - prípady [zd3002rr]
http://datacube.statistics.sk/#!/view/sk/VBD_SK_WIN/zd3002rr/v_zd3002rr_00_00_00_sk
13. Kriminalita podľa základných skupín trestných činov [sk3003rr]
http://datacube.statistics.sk/#!/view/sk/VBD_SK_WIN/sk3003rr/v_sk3003rr_00_00_00_sk
14. Knižnice - knižničné jednotky a výpožičky [ku3006rr]
http://datacube.statistics.sk/#!/view/sk/VBD_SK_WIN/ku3006rr/v_ku3006rr_00_00_00_sk
15. Repertoár a návštevníci divadiel [ku3002rr]
http://datacube.statistics.sk/#!/view/sk/VBD_SK_WIN/ku3002rr/v_ku3002rr_00_00_00_sk
16. Emisie základných znečisťujúcich látok [zp3003rr]
http://datacube.statistics.sk/#!/view/sk/VBD_SK_WIN/zp3003rr/v_zp3003rr_00_00_00_sk
17. Výsledky volieb do NR SR (ideálne všetky od počiatku SR)
<https://volby.statistics.sk/>
18. Atlas rómskych komunít 2019
<https://www.minv.sk/?atlas-romskych-komunit-2019>
stĺpce:
B - okres (bude treba sčítať/spájať jednotlivé mestá do jednotlivých okresov)
F - podiel rómskych obyvateľov (bude treba sčítať/spájať jednotlivé mestá do jednotlivých okresov)
AQ - výskyt komunitných centier
19. <https://www.europeansocialsurvey.org/data/country.html?c=slovakia>
country file za všetky dostupné roky