Aplikácia na spracovanie údajov zo Sčítania obyvateľov domov a bytov z roku 2021

Semestrálna práca AUS1

1. Použité údajové štruktúry

1.1 Sorted Sequence Table

Pre uchovanie načítaných územných jednotiek bola zvolená sekvenčná utriedená tabuľka implementovaná pomocou ArrayListu. Keďže sa prvky do tabuľky pridajú iba na začiatku programu a ďalej sa už tabuľka nemodifikuje, podstatná operácie v tabuľke je nájdi prvok. Vďaka ArrayListu môže ľahko pristupovať k jednotlivým prvkom a keďže sú jej dáta utriedené, môže pomocou bisekcie vyhľadať daný prvok. Zložitosť operácie vyhľadaj je najmenšia spomedzi ostatných tabuliek so zložitosťou O(log(N)).

1.2 Unsorted Sequence Table

Pri filtrovaní územných jednotiek sa dáta, ktoré prešli filtrom, uložia do neutriedenej sekvenčnej tabuľky implementovanej pomocou ArrayListu. Pri filtrovaní nehrá úlohu utriedenie podľa kľúča, takže dáta stačí do tabuľky ukladať postupne. Keďže má táto tabuľka pre vkladanie prvku zložitosť O(1) je najefektívnejšia spomedzi ostatných tabuliek. Taktiež je túto tabuľku možné utriediť podľa zvoleného parametra pokiaľ si užívateľ zvolí triedenie dát. Vďaka ArrayListu je jednoduchý prístup k prvkom a tie sa môžu jednoducho medzi sebou vymieňať.

1.3 Linked List

Zreťazený zoznam bol použitý pri práci s kompozitným filtrom. Keď bude chcieť užívateľ filtrovať, bude mať možnosť vybrať si z viacerých filtrov a tie sa následne zaregistrujú do kompozitného filtra. Keďže sa filtre pridávajú postupne a pri filtrovaní sa prechádza každým jedným filtrom, dôležitá operácia pre zoznam je vlož prvok. Zreťazený zoznam má oproti ArrayListu výhodu v tom, že nemusí alokovať nové pole keď mu dôjde kapacita. Zložitosť operácie vlož je O(1).

1.4 Array

Pre uchovanie informácií o počte obyvateľov v danom veku bolo zvolené pole, keďže rozptyl je vždy od 0 po 100 a prístup k prvkom v poli je jednoduchý so zložitosťou O(1).

2. Popis tried a UML diagramy

2.1 Record

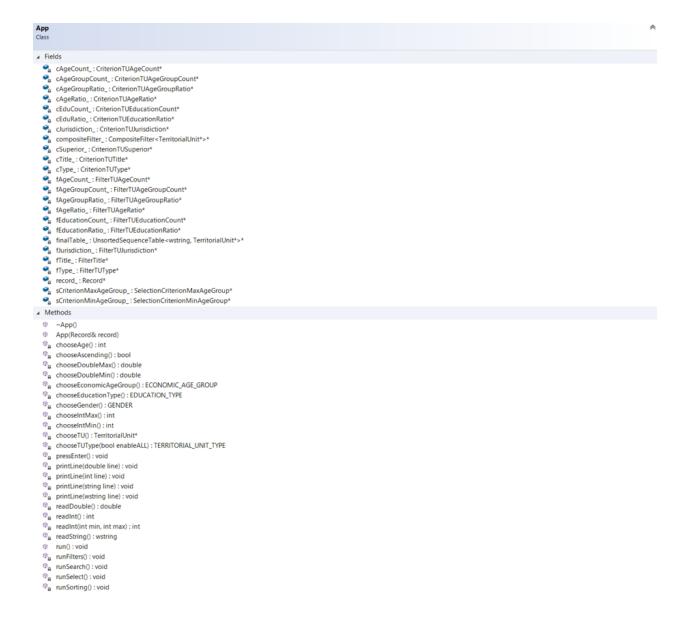
Trieda Record slúži na načítanie územných jednotiek a ich uchovanie. V privátnych poliach si drží informácie o všetkých územných jednotkách v utriedenej sekvenčnej tabuľke. Tiež má

k dispozícii jednu územnú jednotku typu štát reprezentujúcu Slovenskú republiku. Ako posledné si drží prázdnu utriedenú sekvenčnú tabuľku pre prípad, keby užívateľ žiada typ tabuľky, ktorá nevyhovuje žiadnym podmienkam. Hlavnou metódou je metóda loadData() s parametrami reprezentujúcimi jednotlivé názvy súborov. Táto metóda následne spolu s privátnymi pomocnými metódami načíta územné jednotky, vzdelanie a vek a správne ich rozdelí. Ďalšie metódy súžia na prístup k tabuľkám a územným jednotkám. getAllTerritorialUnits() vráti tabuľku všetkých územných jednotiek. getTUTable() vráti tabuľku podľa typu UJ a prípadne nadradenej územnej jednotky. getFilteredTable() preberá ako parameter odkaz na neutriedenú tabuľku a filter a následne vyfiltrované dáta uloží do neutriedenej tabuľky. getSlovakia() vráti Slovensko. Metódou clear() sa vymažú všetky načítané územné jednotky.



2.2 App

Trieda App sa stará o komunikáciu s užívateľom. Ako parameter do konštruktora preberie triedu Record s ktorou ďalej pracuje. Pre začiatok komunikácie s užívateľom je nutné zavolať metódu run() ktorá beží, kým sa užívateľ nerozhodne ukončiť aplikáciu. Pri zvolení konkrétneho druhu operácie sa zavolá jedna z metód runSearch() pre bodové vyhľadávanie, runFilters() pre filtrovanie, runSorting() pre triedenie alebo runSelect() pre výber najlepšieho. Trieda si v privátnych poliach uchováva odkazy na všetky kritériá a filtre s ktorými ďalej pracuje. Taktiež má vytvorenú neutriedenú sekvenčnú tabuľku v ktorej si uchováva vyfiltrované územné jednotky zo záznamu. Ďalej má k dispozícií pomocné privátne metódy na načítanie korektných vstupov a výpis na konzolu.



2.3 Vek

Trieda Age si uchováva informácie o počte obyvateľov v danom veku v závislosti od pohlavia.



2.4 Vzdelanie

Trieda Education si uchováva dáta o vzdelaní. V privátnych poliach má uložený počet ľudí bez ukončeného vzdelania, so základným vzdelaním, učňovským, stredným, vyšším vzdelaním, bez vzdelania a počet nezistených.



2.5 Územné jednotky

2.5.1 Územná jednotka

Trieda TerritorialUnit predstavuje základnú triedu pre všetky územné jednotky. Uchováva si informácie o kóde, oficiálnemu, strednému a skrátenému názvu, poznámke, type, vzdelaní a veku. Tiež si drží informáciu o nadradenej územnej jednotke.



2.5.2 Obec

Trieda Municipality je typ územnej jednotky predstavujúci obec.



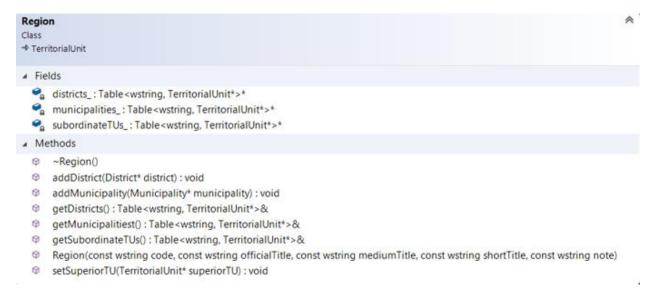
2.5.3 Okres

Trieda District predstavuje okres a okrem dát z územnej jednotky si uchováva ešte aj tabuľku obcí ktoré patria do okresu.



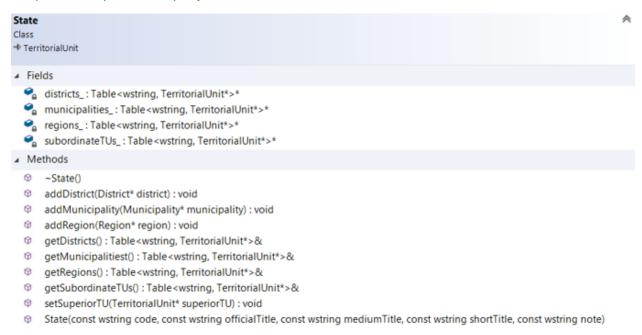
2.5.4 Kraj

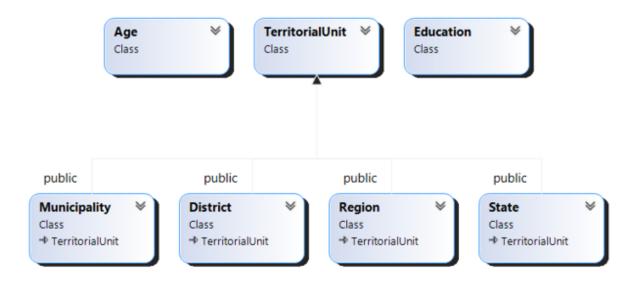
Trieda Region reprezentuje kraj a uchováva si tabuľku obcí, okresov a všetkých podriadených územných jednotiek.



2.5.5 Štát

Trieda State predstavuje štát a uchováva si tabuľku obcí, okresov, krajov a všetkých podriadených územných jednotiek.





2.6 Kritériá

Kritériá slúžia v aplikácii na to aby zo zvolenej územnej jednotky vrátili jednu vlastnosť. Pokiaľ kritérium k ohodnoteniu vlastnosti potrebuje ďalšie atribúty, tieto atribúty sa nastavia v konštruktore ale je ich možné aj zmeniť pomocou verejných metód.

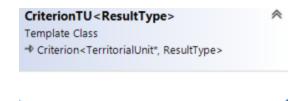
2.6.1 Kritérium

Criterion<ObjectType, ResultType> predstavuje základnú šablónovú triedu pre kritériá. K dispozícii má virtuálnu metódu ohodnoť ktorá vráti požadovanú hodnotu. ObjectType je typ spracovávaného objektu a ResulType je typ návratovej hodnoty.



2.6.2 Kritérium územná jednotka

CriterionTU<ResultType> je základná šablónová trieda pre kritériá pracujúce s územnými jednotkami.



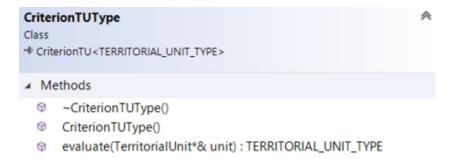
2.6.3 Kritérium UJNázov

CriterionTUTitle vráti názov územnej jednotky.



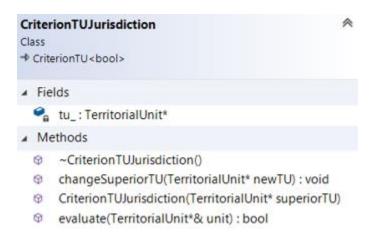
2.6.4 Kritérium UJTyp

CriterionTUType vráti typ územnej jednotky.



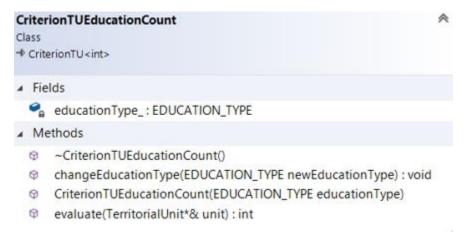
2.6.5 Kritérium UJPríslušnosť

CriterionTUJurisdiction vráti true alebo false podľa toho či územná jednotka patrí zvolenej nadradenej územnej jednotke.



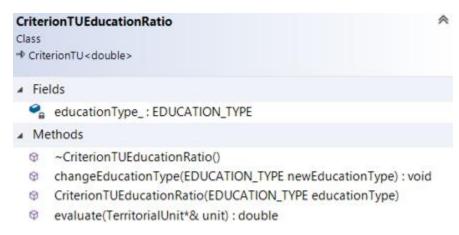
2.6.6 Kritérium UJVzdelaniePočet

CriterionTUEducationCount vráti počet ľudí s daným vzdelaním.



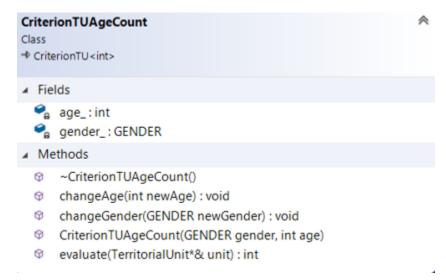
2.6.7 Kritérium UJVzdelaniePodiel

CriterionTUEducationRatio vráti podiel ľudí s daným vzdelaním ku celkovému počtu obyvateľov.



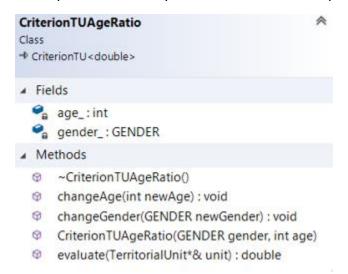
2.6.8 Kritérium UJVekPočet

CriterionTUAgeCount vráti počet ľudí s daným vekom.



2.6.9 Kritérium UJVekPodiel

CriterionTUAgeRatio vráti podiel ľudí s daným vekom ku celkovému počtu obyvateľov.



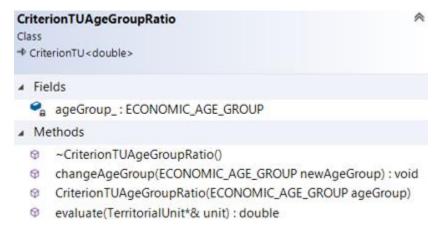
2.6.10 Kritérium UJVekováSkupinaPočet

CriterionTUAgeGroupCount vráti počet ľudí danej vekovej skupiny.



2.6.11 Kritérium UJVekováSkupinaPodiel

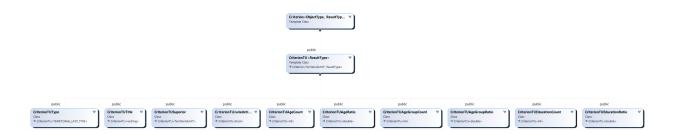
CriterionTUAgeGroupRatio vráti podiel ľudí danej vekovej skupiny ku celkovému počtu obyvateľov.



2.6.12 Kritérium UJNadradená

CriterionTUSuperior vráti nadradenú územnú jednotku.



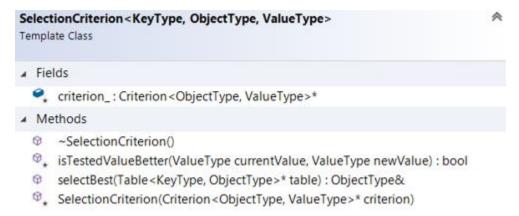


2.7 Výberové kritériá

Výberové kritéria slúžia k tomu, že z prebratej tabuľky vyberú jeden prvok, ktorého požadovaná vlastnosť je najlepšia z pomedzi ostatných prvkov. Pokiaľ výberové kritérium potrebuje k výberu dodatočné atribúty, tie sa mu nastavia v konštruktore ale je ich možné zmeniť pomocou verejných metód.

2.7.1 Výberové kritérium

SelectionCriterion<KeyType, ObjectType, ValueType> je základná trieda pre výberové kritérií. KeyType je typ kľúčov v tabuľke, ObjectType typ spracovávaných objektov a ValueType typ kontrolovanej hodnoty.



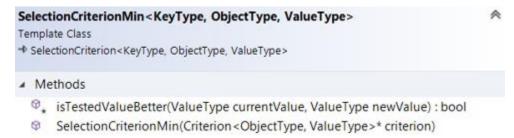
2.7.2 Výberové kritérium zvoľ maximum

SelectionCriterionMax<KeyType, ObjectType, ValueType> je základná šablónová trieda pre výberové kritériá, ktoré vyhľadávajú maximálnu hodnotu. KeyType je typ kľúčov v tabuľke, ObjectType typ spracovávaných objektov a ValueType typ kontrolovanej hodnoty.



2.7.3 Výberové kritérium zvoľ minimum

SelectionCriterionMin<KeyType, ObjectType, ValueType> je základná šablónová trieda pre výberové kritériá, ktoré vyhľadávajú minimálnu hodnotu. KeyType je typ kľúčov v tabuľke, ObjectType typ spracovávaných objektov a ValueType typ kontrolovanej hodnoty.



2.7.4 Výberové kritérium najväčšia veková skupina

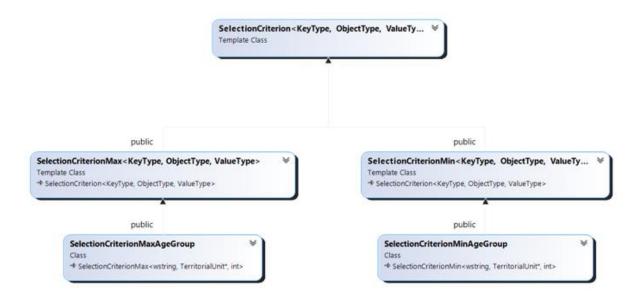
SelectionCriterionMaxAgeGroup vyberie územnú jednotku s najväčším počtom obyvateľov danej vekovej skupiny.



2.7.5 Výberové kritérium najmenšia veková skupina

SelectionCriterionMaxAgeGroup vyberie územnú jednotku s najmenším počtom obyvateľov danej vekovej skupiny





2.8 Filtre

Filtre slúžia k vyfiltrovaniu územných jednotiek na základe danej vlastnosti. Metóda pass() vráti true alebo false na základe toho, či územná jednotka spĺňa danú hodnotu. Každý filter spolupracuje s kritériom z ktorého získava danú vlastnosť územnej jednotky. Hodnoty do filtra a použitého kritéria sa nastavujú v konštruktore ale je ich možné zmeniť pomocou verejných metód. Tiež každý filter má metódu toString() ktorá vráti reťazec popisujúci filter a použité kritérium.

2.8.1 Filter

Filter<ObjectType> je základná šablónová trieda pre filtre. ObjectType je typ objektu ktorý sa bude filtrovať.



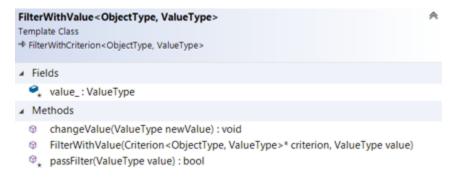
2.8.2 Filter s kritériom

FilterWithCriterion<ObjectType, ValueType> je základná šablónová trieda pre filtre hodnotené pomocou kritéria. ObjectType je typ filtrovaného objektu a ValueType typ kontrolovanej hodnoty.

Templ	WithCriterion < ObjectType, ValueType > late Class	^
+P Filte	er <objecttype></objecttype>	
∡ Fie	elds	
0.	criterion_: Criterion < ObjectType, ValueType>*	
a Me	ethods	
0	~FilterWithCriterion()	
0,	FilterWithCriterion(Criterion <objecttype, valuetype="">* criterion</objecttype,>)
9	pass(ObjectType& object) : bool	
Θ.	passFilter(ValueType value) ; bool	

2.8.3 Filter s hodnotou

FilterWithValue<ObjectType, ValueType> je základná šablónová trieda pre filtre s kritériom kde kontrolovaná hodnota musí mať presnú hodnotu. ObjectType je typ filtrovaného objektu a ValueType typ kontrolovanej hodnoty. Kontrolovaná hodnota sa nastaví v konštruktore ale je možné ju zmeniť pomocou verejnej metódy.



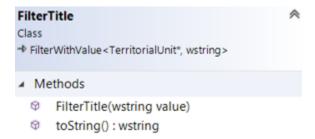
2.8.4 Filter's intervalom

FilterWithInterval<ObjectType, ValueType> je základná šablónová trieda pre filtre s kritériom kde kontrolovaná hodnota musí mať byť v určitom intervale. ObjectType je typ filtrovaného objektu a ValueType typ kontrolovanej hodnoty. Maximum a minimum v intervale sa nastavuje v konštruktore ale je ho možné zmeniť aj verejnými metódami.



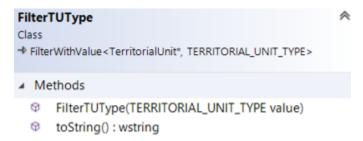
2.8.5 Filter názov

FilterTitle je hodnotový filter porovnávajúci názov územnej jednotky. Filter spolupracuje s kritériom Názov.



2.8.6 Filter UJTyp

FilterUJType je hodnotový filter porovnávajúci typ územnej jednotky. Filter spolupracuje s kritériom UJTyp.



2.8.7 Filter UJPríslušnosť

FilterTUJurisdiction je hodnotový filter porovnávajúci či územná jednotka patrí do zvoleného vyššieho územného celku. Filter spolupracuje s kritériom UJPrislusnost.

2.8.8 Filter UJVzdelaniePočet

FilterTUEducationCount je intervalový filter, ktorý kontroluje či počet obyvateľov s daným vzdelaním patrí do zvoleného intervalu. Filter spolupracuje s kritériom UJVzdelaniePodiel.



2.8.9 Filter UJVzdelaniePodiel

FilterTUEducationRatio je intervalový filter, ktorý kontroluje či podiel obyvateľov s daným vzdelaním patrí do zvoleného intervalu. Filter spolupracuje s kritériom UJVzdelaniePodiel.



2.8.10 Filter UJVekPočet

FilterTUAgeCount je intervalový filter, ktorý kontroluje či počet obyvateľov s daným vekom patrí do zvoleného intervalu. Filter spolupracuje s kritériom UJVekPocet.



2.8.11 Filter UJVekPodiel

FilterTUAgeRatio je intervalový filter, ktorý kontroluje či podiel obyvateľov s daným vekom patrí do zvoleného intervalu. Filter spolupracuje s kritériom UJVekPodiel.



2.8.12 Filter UJVekováSkupinaPočet

FilterTUAgeGroupCount je intervalový filter, ktorý kontroluje či počet obyvateľov v danej vekovej skupine patrí do zvoleného intervalu. Filter spolupracuje s kritériom UJVekováSkupinaPočet.



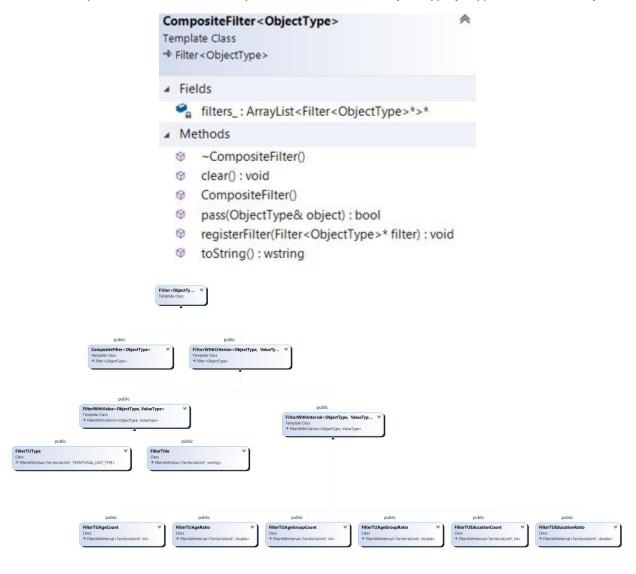
2.8.13 Filter UJVekováSkupinaPodiel

FilterTUAgeGroupCount je intervalový filter, ktorý kontroluje či podiel obyvateľov v danej vekovej skupine patrí do zvoleného intervalu. Filter spolupracuje s kritériom UJVekováSkupinaPodiel.



2.8.14 Kompozitný filter

CompositeFilter<ObjectType> je kompozitný filter do ktorého sa dajú registrovať filtre a pri testovaní prejde každým jedným a zvolený objekt otestuje. Pokiaľ nie je registrovaný žiadny filter tak pri testovaní automaticky vráti hodnotu true. ObjectType je typ filtrovaného objektu.



2.9 Vymenované typy

TERRITORIAL_UNIT_TYPE vymenované typy územných jednotiek.

EDUCATION_TYPE vymenované typy vzdelania.

DATA TYPE vymenované typy načítavaných dát.

ECONOMIC_AGE_GROUP vymenované typy ekonomických vekových skupín.

GENDER vymenované typy pohlaví.

3. Popis zložitostí

3.1 Načítanie dát

Nové údajové štruktúry sa vkladajú do utriedenej sekvenčnej tabuľky podľa názvu pomocou bisekcie so zložitosťou O(log(N)). Avšak keďže ich treba pridať do viacerých tabuliek tak celková zložitoť je O(M*log(N) + log(N)) kde M je počet nadradených územných celkov do ktorých sa územná jednotka musí pridať a posledné log(N) je z dôvodu pridávania všetkých územných jednotiek do spoločnej tabuľky.

Pre korektné vkladanie územných jednotiek do VUC alebo pre vkladanie vzdelania a veku sa používa kód územnej jednotky. Keďže sú však územné jednotky utriedené podľa názvu je potrebné si všetky prejsť so zložitosťou O(N)

3.2 Bodové vyhľadávanie

Pri bodovom vyhľadávaní sa zvolená územná jednotka vyhľadá podľa názvu. Keďže sú územné jednotky vopred utriedené podľa názvu tak celková zložitosť operácie je O(log(N) + k) kde M je počet informácií ktoré sa majú vypísať.

3.3 Filtrovanie

Pre zistenie či územné jednotky spĺňajú podmienky filtra je potrebné ich všetky prejsť a otestovať. Celková zložitosť potom vychádza O(N + k) kde k je počet filtrov ktorými má územná jednotka prejsť. Pri použití filtra UJTyp sa vyberie iba tabuľka s daným typom a pri použití filtra UJPríslušnosť sa vyberie daná tabuľka

iba z vybranej územnej jednotky. Týmto sa zredukuje počet filtrovaných jednotiek čo filtrovanie urýchli.

3.4 Triedenie

Počas triedenia si najskôr užívateľ vyberie filtre a na základe nich získa neutriedenú tabuľku vyhovujúcich prvkov. Pri nezvolení žiadneho filtra sa tiež vykoná filtrovanie s tým, že vyhovujúce budú všetky prvky. Zložitosť filtrovania je opísaná v bode 3.4. Následne sa tabuľka triedi algoritmom QuickSort so zložitosťou $O(N*log(N)) - O(N^2)$.

3.5 Výber najlepších

Pri výbere najlepších si spočiatku užívateľ vyberie filtre ktoré chce aplikovať podobne ako v bodoch 3.3 a 3.4. Následne si zvolí výberové kritérium a to so zložitosťou O(N) prejde celú tabuľku a vráti najlepšiu územnú jednotku.

4. Používateľská príručka

Aplikácia komunikuje s užívateľom pomocou konzoly s jednoduchým výpisom a vstupom. Pri výbere z viacerých možností užívateľa vyzve aby zadal číslo odpovedajúce možnosti. Možnosti výberu sú zvyčajne od 1 po N kde N je počet možností. Pri niektorých výberoch je možné stlačiť 0 čo signalizuje ukončenie výberu. Pre zadávanie parametrov je užívateľ vyzvaný aby zadal korektné hodnoty.

4.1 Hlavné menu

Po spustení aplikácie si užívateľ bude môcť zvoliť z nasledujúcich možností: Bodové vyhľadávanie, Filtrovanie, Triedenie, Výber najlepších alebo Koniec. Možnosť si vyberie stlačením uvedeného čísla. Po vybratí možnosti aplikácie pokračuje k zvolenej úlohe.

4.2 Bodové vyhľadávanie

Pri bodovom vyhľadávaní je užívateľ vyzvaný aby zadal názov územnej jednotky ktorú chce vyhľadať. Následne si zvolí typ vzdelania ktoré chce vypísať a typ ekonomickej vekovej skupiny. Po navolení sa mu vyhľadá územná jednotka a vypíšu sa zvolené informácie. Po stačení klávesy ENTER sa aplikácia vráti do hlavného menu

```
-- Bodove vyhladavanie --
Nazov uzemnej jednotky:
Zilina
Zvolte typ vzdelania:
1. Bez ukonceneho vzdelania
2. Zakladne
3. Ucnovske
4. Stredne
5. Vyssie
6. Vysokoskolske
Bez vzdelania
8. Nezistene
Zvolte ekonomicku vekovu skupinu:

    Predproduktivni

Produktivni
Poproduktivni
Zvolena uzemna jednotka:
Nazov: Zilina
Typ: Obec
Vzdelanie: Vysokoskolske Pocet: 22861
Ekonomicka vekova skupina: Produktivni Pocet: 54975
Vyssie uzemne celky:
Nazov: Okres Zilina
                       Typ: Okres
Nazov: Zilinsky kraj Typ: Kraj
Nazov: Slovenska republika Typ: Stat
Pre pokracovanie stlacte klavesu ENTER
```

4.3 Filtrovanie

Po spustení filtrovania bude užívateľ vyzvaný aby si vybral jeden z uvedených filtrov. Následne je potrebné nastaviť parametre k filtrom. Pokiaľ chce užívateľ zmeniť parametre už zvoleného filtra, môže ho označiť znovu a upraviť. Po ukončení výberu sa zobrazia použité filtre a územné jednotky ktoré filtrami prešli. Po stačení klávesy ENTER sa aplikácia vráti do hlavného menu

```
--- Filtrovanie ---
Vyberte filtre:
1. UJVzdelaniePocet
2. UJVzdelaniePodiel
3. UJTyp
4. UJPrislusnost
5. UJVekPocet
6. UJVekPodiel
7. UJVekovaSkupinaPocet
8. UJVekovaSkupinaPodiel
0. Pokracovat
```

4.4 Triedenie

Na začiatku triedenia si užívateľ navolí kritérium podľa ktorého sa bude triediť. Následne si zvolí spôsob triedenia a ďalej si bude môcť navoliť uvedené filtre v závislosti od použitého kritéria. Následne sa vypíšu utriedené územné jednotky aj s hodnotou kritéria podľa ktorého sa triedili. Po stačení klávesy ENTER sa aplikácia vráti do hlavného menu.

```
--- Triedenie ---
Zvolte kriterium pre triedenie:
1. Nazov
2. UJVzdelaniePocet
3. UJVzdelaniePodiel
4. UJVekPodiel
5. UJVekovaSkupinaPocet
1
Zvolte sposob triedenia:
1. Vzostupne
2. Zostupne
1
Vyberte filtre:
1. UJTyp
2. UJPrislusnost
0. Pokracovat
```

4.5 Výber najlepších

Pri výbere najlepších si užívateľ najskôr zvolí výberové kritérium, ekonomickú vekovú skupinu a bude si môcť vybrať z dvoch filtrov. Po navolení sa mu vypíše najlepšia územná jednotka a hodnota výberového kritéria. Po stačení klávesy ENTER sa aplikácia vráti do hlavného menu.

```
--- Vyber najlepsich ---
Zvolte vyberove kriterium:

    NajvacsiaVekovaSkupina

NajmensiaVekovaSkupina
Zvolte ekonomicku vekovu skupinu:

    Predproduktivni

Produktivni
Poproduktivni
Vyberte filtre:

    UJTyp

UJPrislusnost
Pokracovat
Najlepsia uzemna jednotka:
Nazov: Slovenska republika
Typ: Stat
Vekova skupina: Produktivni, Pocet: 3652679
```