

SVEUČILIŠTE U ZAGREBU

FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

# Projekt kolegija Ekspertni sustavi

Autor: Roko Krstulović

Zagreb, siječanj 2017.

## Sadržaj

1. Uvod .....	1
2. Baza znanja sustava.....	1
2.1. Datoteke <i>Sportovi.txt</i> , <i>Pitanja.txt</i> , <i>Vjerojatnosti.txt</i> .....	1
2.2. Podaci u bazi podataka .....	2
3. Funkcioniranje ekspertnog sustava.....	2
3.1. Vjerojatnosti sportova .....	3
3.2. Odabir pitanja .....	4
4. Implementacija i pokretanje ekspertnog sustava .....	4
4.1. Implementacija ekspertnog sustava.....	4
4.2. Pokretanje ekspertnog sustava.....	5
5. Zaključak .....	6

## 1. Uvod

Kao zadatak ovog projekta odabrana je implementacija ekspertnog sustava koji korisniku pomaže u odabiru sporta. Ekspertni sustav koristi bazu znanja u kojoj su pohranjene osobine sportaša u obliku vjerojatnosti te na temelju njih postavlja pitanja „pametnim“ redoslijedom kako bi korisnik što prije došao do sporta koji njemu odgovara. U svakom trenutku ekspertni sustav prikazuje *broj\_najvjerojatnijih* (default je tri) sportova, te na temelju odgovora na pitanja ih mijenja (ako je potrebno).

## 2. Baza znanja sustava

Ekspertni sustav za pomoć u odabiru sporta koristi bazu znanja u kojoj su pohranjene osobine sportaša u obliku vjerojatnosti. U ovom će poglavlju biti objašnjena datoteke po kojima su pohranjeni podaci te sami podaci.

### 2.1. Datoteke *Sportovi.txt*, *Pitanja.txt*, *Vjerojatnosti.txt*

Baza znanja je pohranjena unutar tri datoteke: *Sportovi.txt*, *Pitanja.txt* i *Vjerojatnosti.txt*.

Svaki red datoteke *Sportovi.txt* prikazuje definiciju jednog sporta. Definicija sporta izgleda ovako:

<id\_sporta> <ime\_sporta> <brojnost\_sporta> <slika\_sporta>

<id\_sporta> je primarni ključ sporta (jedinstveno predstavlja sport) te se koristi unutar datoteke *Vjerojatnosti.txt*. <ime\_sporta> predstavlja ime sporta. <brojnost\_sporta> predstavlja brojnost sportaša određenog sporta. Brojnost sportaša se koristi za računanje apriorne vjerojatnosti sporta. <slika\_sporta> je putanja do slike koja će biti prikazana u slučaju da je sport među *broj\_najvjerojatnijih* najvjerojatnija za korisnika.

Unutar datoteke *Pitanja.txt* definirana su pitanja i ponuđeni odgovori na ta pitanja. Svako pitanje ima dva ili više odgovora. Za svaku kombinaciju pitanje-odgovor potrebno je moći definirati vjerojatnost da je sportaš odgovorio baš taj odgovor. Te vjerojatnosti definirane su u datoteci *Vjerojatnosti.txt* te će format datoteke biti kasnije objašnjen. Zbroj vjerojatnosti odgovora za jedno pitanje iznosi 100% te je potrebno definirati vjerojatnosti za sve odgovore osim jednog (kada gledamo samo jedno pitanje). Datoteka *Pitanja.txt* pridaje svim odgovorima osim jednog po pitanju primarni ključ koji se koristi u datoteci *Vjerojatnosti.txt* gdje se tim odgovorima pridodaju vjerojatnosti (npr. za pitanje „Jeste li visoka osoba?“ S odgovorima „Da“ i „Ne“ postojati će jedan primarni ključ za odgovor „Da“ a vjerojatnost za odgovor će biti izračunata). Pitanje s N odgovora je definirano ovako:

<id\_prvog\_odgovora> <tekst\_pitanja> [<prvi\_odgovor>; <drugi\_odgovor>; ... ; <n-ti\_odgovor>]

<id\_drugog\_odgovora>...

<id\_N-1-og\_odgovora>

<id\_prvog\_odgovora>, <id\_drugog\_odgovora>, ..., <id\_N-1-og\_odgovora> predstavljaju primarne ključeve odgovora od 1 do N-1. <tekst\_pitanja> sadrži pitanje zajedno s upitnikom ako je potreban. <prvi\_odgovor>, <drugi\_odgovor>, ..., <n-ti\_odgovor> predstavljaju tekstove odgovora.

Datoteka *Vjerojatnosti.txt* sadrži vjerojatnosti da je sportaš iz sporta S odgovorio na odgovor <id\_odgovora> za sve odgovore osim jednog po pitanju. Vjerojatnost je definirana ovako:

$$P(\langle \text{id\_odgovora} \rangle | \langle \text{id\_sporta} \rangle) = \langle \text{vjerojatnost} \rangle$$

<id\_odgovora> predstavlja primarni ključ za kombinaciju pitanje-odgovor.

<id\_sporta> predstavlja primarni ključ sporta.

<vjerojatnost> predstavlja vjerojatnost da će sportaš iz <id\_sporta> odgovoriti <id\_odgovora> u postocima.

Datoteka također dopušta prazne retke i komentare koji počinju s znakom # koji se ignoriraju prilikom parsiranja.

## 2.2. Podaci u bazi podataka

Baza podataka sadrži podatke o osobinama sportaša u obliku vjerojatnosti. U bazi se nalaze 31 različita sporta, 18 pitanja i 39 ponuđenih odgovora. Dodatno, u bazi se nalazi 651 vjerojatnost koja povezuje sportaše (i njihove sportove) te njihove osobine (odnosno odgovore na pitanja). Za funkcioniranje aplikacije potrebno je dodati 558 vjerojatnosti koje se izračunavaju koristeći činjenicu da suma vjerojatnosti svih odgovora po jednom pitanju iznosi 100%. Podaci su dobiveni ispitivanjem jako malog broja sportskih eksperata, te se zasigurno mogu i moraju poboljšati anketiranjem stvarnih profesionalnih i amaterskih sportaša.

## 3. Funkcioniranje ekspertnog sustava

U ovom odlomku biti će objašnjene ideje koje stoje u pozadini same implementacije ekspertnog sustava. U poglavlju 3.1. biti će objašnjeno računanje vjerojatnosti sportova, dok će u poglavlju 3.2. biti objašnjen odabir pitanja koja će se postavljati.

### 3.1. Vjerojatnosti sportova

Sustav u svakom trenutku sadrži vjerojatnosti o tome koliko klijent odgovara svakom od sportova na temelju svih odgovora koje je do sad dao. Vjerojatnost koja nas zanima matematički se prikazuje ovako:

$$P(sport|odgovor1, odgovor2, \dots, odgovorN)$$

i predstavlja vjerojatnost da je sport za osobu koja je dala odgovore od 1, do N. Uz pomoć *Bayesovog*<sup>1</sup> pravila dobivamo:

$$P(sport|odgovor1, \dots, odgovorN) = \frac{P(odgovor1, \dots, odgovorN|sport) * P(sport)}{P(odgovor1, \dots, odgovorN)}$$

Pošto nas zanima koji od sportova ima najveću vjerojatnost, a ne sama vjerojatnost, te pošto vjerojatnost svakog sporta ima istu iznos u nazivniku, nazivnik možemo zanemariti, odnosno imamo:

$$P(sport|odgovor1, \dots, odgovorN) \propto P(odgovor1, \dots, odgovorN|sport) * P(sport)$$

Radi jednostavnosti uvesti ćemo pretpostavku da su odgovori za različita pitanja međusobno nezavisni te možemo pisati:

$$P(sport|odgovor1, \dots, odgovorN) \propto P(odgovor1|sport) * \dots * P(odgovorN|sport) * P(sport)$$

Pošto s desne strane može biti jako puno vjerojatnosti koje će ako se prikazuju u intervalu [0, 1] jako brzo iščeznuti, ili ako su u intervalu [0, 100] eksplodirati iskoristiti ćemo jedan trik. Naime, nećemo promatrati same vjerojatnosti, već ćemo gledati logaritam vjerojatnosti te koristeći logaritamske identitete<sup>2</sup> zbrajati manje brojeve. Pošto je logaritam monotono rastuća funkcija vrijedi:

$$\forall x_1, x_2 \mid x_1 > x_2 \Rightarrow \log(x_1) > \log(x_2)$$

Nakon svih promjena, sport je vjerojatniji što ima veći iznos sljedeće funkcije:

$$\log(P(odgovor1|sport)) + \dots + \log(P(odgovorN|sport)) + \log(P(sport))$$

Nakon što korisnik odabere odgovor, za svaki se sport iz baze znanje uzima vjerojatnost odgovora  $P(<id\_odgovora>|<sporta>)$  koja se zatim logaritmiri i pridodaje trenutnoj „vjerojatnosti“. Bitno je također naglasiti da  $P(<id\_sporta>)$  nije direktno spremljena u bazi znanja, već se računa iz brojnosti sporta:

$$P(sport) = \frac{N(sport)}{\sum_{s \in sportovi} N(s)}$$

---

<sup>1</sup> [Bayesovo pravilo](#)

<sup>2</sup> [Logaritamski identiteti](#)

### 3.2. Odabir pitanja

Sustav „pametno“ odabire pitanje tako da što bolje odijeli najvjerojatnije sportove. Prilikom odabira pitanja sustav prođe kroz sva nekorisćena pitanja, te računa dobrotu pitanja i odabire onu s najvećom dobrotom. Dobrota pitanja je definirana razlikom vjerojatnosti svih odgovora tog pitanja za svaki od susjeda najvjerojatnijih sportova. Na taj način početna pitanja su najrelevantnija, a pitanja koja su pri kraju često nemaju skoro nikakav utjecaj. Jednom kad su postavljena sva pitanja, sustav prikaže poruku „Nema više pitanja!“.

## 4. Implementacija i pokretanje ekspertnog sustava

U ovom će se ulomku govoriti o samoj implementaciji ekspertnog sustava te kako ga pokrenuti.

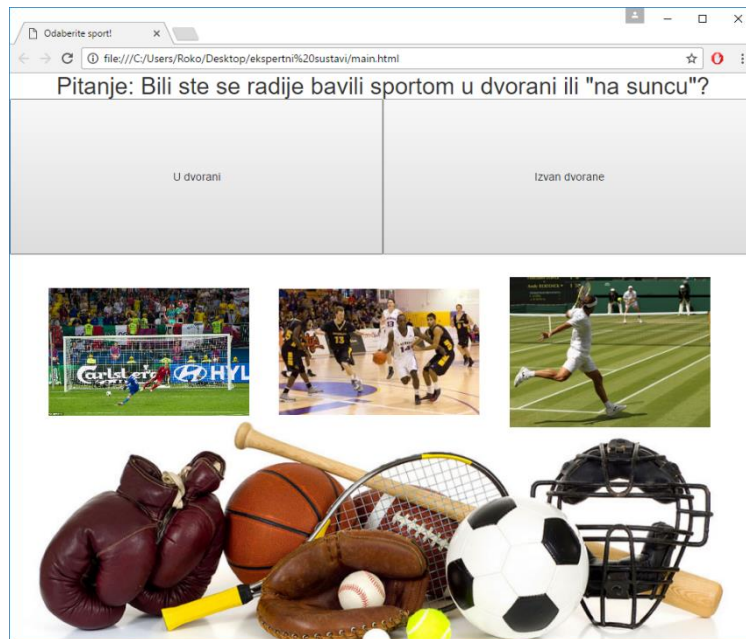
### 4.1. Implementacija ekspertnog sustava

Ekspertni sustav napravljen je kao web aplikacija čiji cijeli kôd se nalazi na klijentskoj strani. Baza znanja se nalazi unutar tri datoteke koje bi idealno bile pohranjene na poslužitelju, no trenutno se također nalaze na klijentu. Sustav je implementiran korištenjem sljedećih tehnologija:

- HTML
- CSS
- Javascript
- AngularJS

Izgled (*front-end*) aplikacije sastoji se od tri dijela: pitanja, odgovora i najvjerojatnijih sportova, te je prikazan je na Slici 1.

Pozadina (*back-end*) aplikacije sastoji se od jednog upravljača AngularJS-a. Upravljač čita podatke iz datoteka u kojima je pohranjena baza znanja iz kojih kreira modele. Upravljač ujedno kreira funkcije za postavljanje i odabir pitanja te za odgovor na pitanje. Aplikacija iskorištava dvostruko vezivanje (engl. *two way binding*) AngularJS-a tako da promjena modela automatski mijenja prikaz web aplikacije. Na taj način promjena najvjerojatnijih sportova unutar modela mijenja slike u izgledu aplikacije, promjena modela pitanja mijenja pitanje i odgovore prikazane na ekranu.



Slika 1. Izgled ekspertnog sustava

#### 4.2. Pokretanje ekspertnog sustava

Uz dokumentaciju priložena je i ZIP datoteka. Ona se sastoji od sljedećih za sustav neophodnih datoteka i mapa:

- Datoteke *main.html* – izvorni kod klijentske strane
- Datoteke *Pitanja.txt* – dio baze znanja vezane za pitanja
- Datoteke *Sportovi.txt* – dio baze znanja vezane za sportove
- Datoteke *Vjerojatnosti.txt* – dio baze znanja vezane za vjerojatnosti
- Mape *img* koja se sastoji od 32 slike različitih sportova – slike koje se prikazuju za najvjerojatnije sportove

Dodatno priložena ZIP datoteka može imati:

- Datoteke *script.py* – python datoteka koja uz pomoć datoteka *Pitanja.txt* i *Vjerojatnosti.txt* generira datoteku predložka za vjerojatnosti *Vjerojatnosti\_TMP.txt*
- Datoteku *chrome.cmd* koja sadrži komandu za pokretanje aplikacije *Google Chrome* u kojem dopušta čitanje iz datoteka, te otvara *main.html* stranicu (za Windowse).

Za pokretanje sustava potrebno je raspakirati priloženu ZIP datoteku te napraviti dvije izmjene:

- U *main.html* datoteci izmijeniti vrijednosti varijabla *putanja\_do\_pitanja*, *putanja\_do\_sportova* i *putanja\_do\_vjerojatnosti* na putanje do datoteka: *Pitanja.txt*, *Sportovi.txt* i *Vjerojatnosti.txt*
- (Opcionalno) U datoteci *chrome.cmd* promijeniti putanju do *chrome.exe* (ako je potrebno) te putanju do *main.html* datoteke.

Aplikacija se pokreće tako da se otvori *main.html* preglednikom kojemu je omogućeno čitanje iz datoteka (ako imate datoteku *chrome.cmd* i napravili ste opcionalne izmjene možete samo pokrenuti tu datoteku kao naredbu).

## 5. Zaključak

Kroz ovaj projekt napravljen je ekspertni sustav za odabir sporta kao web aplikacija. Uz poznato Bayesovo pravilo i uz dobru bazu znanja (dobivenu kroz jednostavne ankete) pokazali smo da nije problem napraviti sustav koji će relativno dobro moći preporučiti sport korisniku preko samo nekoliko jednostavnih pitanja. Najteži dio ovog projekta (kao i u većini ekspertnih sustava) bila je izrada baze znanja.

Ekspertni sustav podsjeća na sustav utemeljen na pravilima. Za razliku od sustava utemeljenih na pravilima ovaj ekspertni sustav računa vjerojatnosti i na taj način se međusobno sportovi ne isključuju. Sustav, ustvari, koristi najpliću bayesovu mrežu te korištenjem bayesovog pravila efikasno računa tražene vjerojatnosti.