SVEUČILIŠTE U ZAGREBU

**FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA**

Projekt kolegija Ekspertni sustavi

Autor: Roko Krstulović

Zagreb, siječanj 2014.

Sadržaj

[1. Uvod 3](#_Toc472464464)

[2. Baza znanja sustava 3](#_Toc472464465)

[2.1. Datoteke *Sportovi.txt*, *Pitanja.txt*, *Vjerojatnosti.txt* 3](#_Toc472464466)

[2.2. Podaci u bazi podataka 4](#_Toc472464467)

[3. Implementacija ekspertnog sustava 4](#_Toc472464468)

[3.1. Vjerojatnosti sportova 4](#_Toc472464469)

# 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 bolje razdijelio najvjerojatnije sportove. U svakom trenutku ekspertni sustav prikazuje tri najvjerojatnija sporta, te na temelju odgovora na pitanja ih mijenja (ako je potrebno).

# 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.

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

Ekspertni sustav za pomoć u odabiru sporta koristi bazu znanja u kojoj su pohranjene osobine sportaša u obliku vjerojatnosti. Baza 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 (jedinstveno predstavlja sport) sporta te se koristi unutar datoteke *Vjerojatnosti.txt*. <ime\_sporta> predstavlja ime sporta. <brojnost\_sporta> 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 tri 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 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“ biti će 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> <text\_pitanja> [<prvi\_odgovor>; <drugi\_odgovor>; … ; <n-ti\_odgovor>]

<id\_drugog\_odgovora>…

<id\_n-1-og\_odogovora>

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

P(<id\_odgovora>|<id\_sporta>)=<vjerojatnost>

<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.

## Podaci u bazi podataka

Baza podatka 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 dodanih 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.

# Implementacija ekspertnog sustava

U ovom odlomku biti će objašnjene ideje i sama implementacija ekspertnog sustava. U poglavlju 3.1. biti će objašnjeno računanje vjerojatnosti sportova, dok će u poglavlju 3.1. biti objašnjen odabir pitanja koja će se postavljati.

## Vjerojatnosti sportova

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

i predstavlja vjerojatnost da je sport za osobu koja je odgovorila s odgovorima od 1, do N. Uz pomoć *Bayesovog[[1]](#footnote-1)* pravila dobivamo:

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

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

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 nego ćemo gledati logaritam vjerojatnosti te koristeći logaritamske identitete[[2]](#footnote-2) ćemo zbrajati manje brojeve. Pošto je logaritam monotono rastuća funkcija vrijedi:

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

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

Zaključak: sliči na ES pravila, bayesove mreže

1. [Bayesovo pravilo](https://en.wikipedia.org/wiki/Bayes'_theorem#cite_ref-2) [↑](#footnote-ref-1)
2. [Logaritamski identiteti](https://en.wikipedia.org/wiki/List_of_logarithmic_identities#Using_simpler_operations) [↑](#footnote-ref-2)