

TARTU ÜLIKOOL
Arvutiteaduse instituut
Informaatika õppekava

Martin Reinsalu

Jalgpalliuudiste koostaja „Newerator“ algversioon

Bakalaureusetöö (9 EAP)

Juhendaja: Krista Liin

Tartu 2017

Jalgpalliuudiste koostaja „Newerator“ algversioon

Lühikokkuvõte:

Bakalaureusetöö eesmärk oli luua programm, mis on võimeline genereerima eestikeelseid jalgpalli tulemusuudiseid Euroopa viie suurima kõrgliiga – Inglismaa, Saksamaa, Hispaania, Itaalia, Prantsusmaa – kohta. Valminud programm aitab lihtsustada ajakirjaniku tööd jalgpalliuudise koostamisel ja tuua tavalugejani valitud mängu uudis, ilma et peaks seda mõnelt kindlalt veebilehelt otsima.

Võtmesõnad:

Newerator, jalgpall, ajakirjandus, teksti genereerimine

CERCS: P170 – Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine

The original version of soccer news generator „Newerator“

Abstract:

The aim of this thesis was to create a program, which is able to generate Estonian language soccer news. The program helps to simplify work of a sports journalist and allows a common user to get news of a certain game, without searching the results from websites.

Keywords:

Newerator, soccer, journalism, text generation

CERCS: P170 – Computer science, numerical analysis, systems, control

Sisukord

Sissejuhatus	4
1. Idee programmi loomisest	5
2. Programmi praktiline vajadus	8
3. Sarnased lahendused	9
4. Programmi kirjeldus	10
4.1. Programmis kasutatavad andmed	10
4.2. Programmi kasutamine ja funktsionaalsused	13
5. Teadaolevad probleemid ja võimalikud lahendused	17
6. Visioon programmi edasi arendamiseks	18
Kokkuvõte	19
Kasutatud kirjandus	20
Litsents	22

Sissejuhatus

Sport on tänapäeva meelelahutusliku poole üks lahutamatu osasid. Igapäevaselt toimuvad võistlused nii individuaalaladel ja meeskondlikel aladel maailma eri paikades. Ala populaarsuse määrab ära huviliste arv ja sellest tingituna on meedia poolne kajastus teatud alade suhtes suurem.

Tuntud infoportaali Worldatase [1] viimaste (17.03.2017 seisuga) andmete põhjal on eri riikides vaadatuim spordiala jalgpall [2]. Võimatu on teada saada palju on maailmas jalgpallureid ja klubisid, sest see arv on pidevalt muutumises. Jalgpalli populaarsusele annab suure tõuke kaugel ajalugu, mis sai alguse 1863. aastal Inglismaal [3]. Enam kui 150 aasta jooksul on välja kujunenud suurklubid ja konkurents omale parimaid mängijad palgata on väga tihe. Summad tänapäevases jalgpallis on nii suured, et inimesed kutsuvad tihti klubivahetuse üleminekutasusid kui ka mängijate palganumbreid kosmilisteks.

Jalgpallihuvilistele pakuvad huvi kõmulise poole kõrval rohkem tulemusuudised, mis annavad ülevaate toimunud mängust või mingi kindla liiga ühe vooru mängude tulemustest. Seeläbi saab lugeja teada, milline oli võitjameeskond, kas mäng jäi viiki või kuidas kohtumine sõltuvalt mängitavast klubisarjast kulmineerus. Jalgpalliteemalisi uudiseid on erinevaid ning tihti koosnebki info vaid faktidest: väravalööjate nimed, klubide nimed, mängu tulemus.

Käesoleva bakalaureusetöö eesmärk on luua programm, mis internetist kogutud Saksamaa, Inglismaa, Hispaania, Prantsusmaa ja Itaalia jalgpalli kõrgliigades aset leidnud mängude info põhjal loob grammatiliselt korrektse eestikeelse tulemusuudise.

Töö käigus on uuritud eestikeelsetes uudisteportaalides leiduvaid jalgpalliteemalisi uudiseid. Praktilises pooles kasutatakse programmeerimiskeelt Java, versiooniga 8.

Töö jaguneb neljaks osaks. Esmalt on välja toodud põhjus, millest idee antud teema suhtes tekkis ja milline oli taolise programmi praktiline vajadus. Edasi kirjeldatakse programmi kasutamist ja selle toimimist. Viimasena tuuakse välja sarnaste lahenduste olemasolu ja töö autori poolne visioon valminud lahendust edasi arendada.

1. Idee programmi loomisest

Lõputöö autori hobiks on töötada jalgpalliteemalisi uudiseid kajastavas portaalis soccernet.ee [4]. Ligi kolm aastat kogemust ajakirjandusliku poolega tegeledes on andnud hea ülevaate headest ja halbadest külgedest uudiste koostamisel ja vajaliku info sidumiseks tervikuks lugejaskonnale.

Autor väidab, et protsess keskmine, st 2-3 lõigu ja kuni kümne lause pikkuse, uudise koostamisel on liigselt aeganõudev. Kogemuste põhjal saab näitena välja tuua järgmised sammud, mis on vajalikud tulemusuudise koostamiseks:

1. Avada veebibrauser ja minna veebilehele, kus on soovitud liiga mängude statistika olemas.
2. Sisestada andmebaasi uued tulemused, mille põhjal uuendatakse tabeliseis.
3. Kokku on lepitud, et tuleks sisestada teksti aknasse mängu tulemus koos väravalööjate, omavärava lööjate, penaltil eksinud mängijate ja punase kaardi teeninud mängijate nimedega, mille juures on ka vastava sündmuse minut. Seejärel vormistada tulemuse tekst paksult (inglise k. *bold*) ja penaltil eksimiste ning punase kaardi teenimise olukorrad alla joonitult.
4. Koostada pealkiri, mis oleks tabav ja meelitaks uudist lugema.
5. Koostada sissejuhatav lõik ja vormistada see paksult.
6. Koostada tekst, mis võtab kokku mängu tulemuse.
7. Lisada uudisega kooskõlas olev pilt, näiteks väravalööja.
8. Kokku on lepitud, et tuleks kopeerida viimati uuendatud tabeliseis teksti lõppu
9. Sisestada märksõnad, milleks on väravalööjad, klubide nimed ja riigi nimi, milles vaatluse all olev liiga tegutseb.
10. Kontrollida üle grammatiline pool.
11. Teha kinnitused uudise koostamise keskkonnas, kus tuleb valida põhiteema ja kinnitada uudise avalikustamine.

Uudis valmib minimaalselt 15 minutiga, millest teksti kirjutamisele kulub hinnanguliselt kümme minutit. Võib ette tulla inimlikke eksimusi, kus pärast uudise avalikustamist ja teksti lugemist tuleks sõnastust muuta, on jäänud mõni kirjaviga sisse või mõni oluline fakt on uudisesse kirjutamata jäänud. Sellisel juhul tuleks uudise loomise koht uuesti avada, teha vastavad parandused ja kinnitada muudatused.

Lõputöö autor on uurinud erinevate portaalide poolt koostatud jalgpalli tulemusuudiseid, mis vastavad kriteeriumile anda lugejale edasi info mingist kindlast tulemusest. Erinevate uudiste põhjal on märgata ühisosa, kus enamus uudiseid ei sisalda visuaalset kirjeldust toimunud mängust, vaid edasi on antud faktid.

Järgnevalt on välja toodud paar näidet eelmainitud tähelepaneku kohta.

Messi tõi Barcelonale kolm punkti ja kullavõimalus jäi alles



Ats Kuldkapp

JAGA FACEBOOKIS

KOMMENTEERI

PRINDI ARTIKKEL

JAGA ARTIKLIT

16. aprill 2017 14:04



Lionel Messi. | FOTO: Xinhua/Xinhua/Sipa USA/Scanpix

Hispaania jalgpalli meistriliigas olid eile õhtul võidukad tabeli kaks esimest Madridi Real ja Barcelona.

Real alistas võõrsil Gijoni 3:2. Madridi klubile tõi kolm tähtsat punkti Isco tabamus 90. minutil, kes oli täpne ka 17. minutil. Lisaks skooris Reali poolelt 59. minutil Alvaro Morata.

Barcelona oli kodus samuti 3:2 üle Real Sociedadist, kusjuures kõik kohtumise väravad löödi esimesel poolajal. Barcelona liider Lionel Messi tegi skoori 17. ja 37. minutil, lisaks andis ta 44. minutil väravasöödu Alcacer Pacole. 42. minutil sai Barcelona kaitsja Samuel Umititi hakkama omaväravaga ja teise Real Sociedadadi tabamuse eest hoolitses 45. minutil Xabi Prieto.

Realil on tabeliliidrina koos 75 punkti, Barcelonal on punkte 72. Reali kasuks mängib ka fakt, et neil on pidada jäänud seitse kohtumist, Barcelonal kuus. Nädala pärast kohtutakse Madridis El Clasico raames omavahel. Kolmandal kohal on Madridi Atletico 65 punktiga.

Joonis 1. Tõmmis postimees.ee uudisest [5] mitme tulemuse puhul.



Angel Di Maria **Autor:** Reuters/Scanpix

Prantsusmaa jalgpalli kõrgliiga reedeses ainsas kohtumises oli Pariisi Saint-Germain võõrsil 2:0 üle Angersist.

Mõlema pealinna klubi värava eest hoolitses argentiinlasest ääremängija Angel Di Maria, kes oli täpne 28. ja 84. minutil.

PSG-l on nüüd 32 mänguga koos 74 punkti, sama palju, kui ühe mängu vähem pidanud Monacol. Samuti 32 mängu pidanud Nice'il on 70 punkti.

Monaco võõrustab täna 19. kohal olevat Dijoni, Nice 17. kohal olevat Nancyd. 30. aprillil sõidab PSG külla Nice'ile.

Toimetaja: Anders Nõmm

Joonis 2. Tõmmis err.ee uudisest [6] ainsa tulemuse puhul.

Lõputöö autor on järeldanud, et taolisi uudiseid oleks võimalik koostada ilma inimese mõttetöö kaasabitajaga ja oluliselt väiksema ajakuluga. Selleks oleks vaja luua vastav tarkvara (programm), mis saab veebist kätte vajaliku faktilise info, suudab koostada grammatiliselt korrektseid eestikeelseid lauseid ja siduda need terviklikuks tekstiks.

2. Programmi praktiline vajadus

Programmi kasutajad jagunevad kaheks: tavakasutaja ja ajakirjanik. Kummagi kasutaja puhul pole programmi kasutamine mingil viisil piiratud, aga genereeritava teksti kasutamine võib erineda.

Tavakasutaja jaoks on tähtis lugeda uudist ja saada kätte peamine info tulemusest: kumb meeskond võitis ja kes olid väravalööjad. Ilma Neweratorita on kasutajal võimalus teha valik erinevate portaalide hulgast, aga vajalik info võib olemas olla mõnes teises uudisteportaalil või pole teadaolevalt toimunud mängu tulemusuudist veel olemas. Võrreldes Neweratoriga tuleb sellisel juhul ajakulu suurem.

Neweratori efektiivsus avaldub ülevaatlikkuses, aja kokkuhoius ja grammatilises korrektsuses. Programm suudab mingi kindla tulemusuudise koostada maksimaalselt 40 sekundi jooksul ja väiksema toimunud mängude arvu korral on see aeg lühem. Neweratori abil genereeritud uudis annab põhjaliku ülevaate mängust, peale visuaalse kirjelduse, milles on olemas mängu peamine info õigekirjavigadeta eesti keeles.

Kui mingis kindlas uudisteportaalil pole olemas kõiki neid tulemusuudiseid, mille kohta kasutaja soovib infot saada, peab ta orienteeruma mitme erineva portaali vahel ja seeläbi kasvab ajakulu. Sarnast olukorda ei saa tekkida Neweratori puhul, sest programmi valikute hulgas olevate liigade korral on alati olemas võimalus genereerida kõige hiljutisema tulemuse põhjal uudis.

Ajakirjaniku poolt vaadatuna on Newerator eelkõige töövahend. Genereeritud uudist saab vajadusel täiendada ja samas teha vähese ajakuluga valmis uudise lugejaskonnale. Samuti on võimalik kiiresti teada saada ilma veebibrauserit avamata liigade tabeliseisud, tulemused, eesseisvad mängud ja toimund mängu statistiline pool ning nende andmete põhjal hakata uudist koostama.

3. Sarnased lahendused

Neweratorile sarnaseid lahendusi on loodud teisigi. Nendest üks peamine on StatsMonkey [11], mis on saanud edasiarendamise teel praegu uueks nimeks Quill [12]. Tegemist on tarkvaraga, mis on orienteeritud pesapalli ja suudab automaatselt kirjutada lugusid selle spordialale kohta.

Sarnaselt Neweratorile oli StatsMoneky eesmärk vähendada inimese osalust spordiuudiste kirjutamisel, sest uudiste struktuur on tihti sama ning seetõttu pole korduv ajakulu kuidagi põhjendatud. StatsMonkey annab lugejale edasi faktilise poole mänguhetkede (inglise k. *play-by-play*) põhjal [13].

Martin Molina on oma uurimuses [14] juhtinud tähelepanu asjaolule, et kasulik oleks kasutada tehisintellekti valdkonda kuuluvaid lähenemisi uudiste loomiseks, mis kasutavad olemasolevaid andmeid ja ei tegele visuaalse poolega. Oma töös Molina uurib süsteemi VSAIH, mis hüdroloogiliste andmete põhjal loob inimesele arusaadavaid uudiseid.

Siiani pole ühtegi sarnast lahendust, mis oleks eesti keelega kuidagi seotud. Seega on käesoleva lõputöö käigus loodud programm „Newerator“ esimene taoline lahendus.

4. Programmi kirjeldus

Antud lahendus on loodud programmeerimiskeele Java abil, vajalik on versioon 8. Kasutuses on järgnevad teek Selenium 3.4 [7]. Toimiv internetiühendus on tingimata vajalik.

Lõputöö käigus valminud programm opereerib Euroopa viie suurima liigaga. Nendeks on Inglismaa kõrgliiga ehk Premier League, Saksamaa kõrgliiga ehk Bundesliga, Hispaania kõrgliiga ehk La Liga, Itaalia kõrgliiga ehk Serie A ja Prantsusmaa kõrgliiga ehk Ligue 1.

Saksamaa kõrgliigas osaleb 18 meeskonda ja mängitakse 34 vooru ühel hooajal. Ülejäänud liigades osaleb 20 meeskonda ja mängitakse 38 vooru.

4.1. Programmis kasutatavad andmed

Programmi tööprotsessi käigus kogutud info pärineb järgmistelt veebilehtedelt:

- goal.com [8],
- skysports.com [9],
- footballsquads.co.uk [10].

Erinevate liigade tulemuste ja järgmiste mängude info pärineb goal.com-i lehelt. Samuti mingi kindla mängu kohta käiv statistika kummagi meeskonna suhtes, mis koosneb:

- pallivaldamise protsendist,
- edukalt teise mängijani jõudnud söötude arvust,
- väravale tehtud pealelöökidest arvust,
- väravale tehtud pealelöökidest arvust, mille tulemusena jõudis pall värava raamistikku, kuid ei pruukinud jõuda väravasse,

Uudisteportaali skysports.com-i abil saab programm kätte kõige värskema tabeliseisu, mille komponentideks on iga meeskonna lõikes:

- koht
- meeskonna nimetus,
- võitude arv,
- viikide arv,
- kaotuste arv,

- löödud väravate arv,
- oma väravasse lastud väravate arv,
- löödud ja sisse lastud väravate vahe,
- kogutud punktid.

Veebilehelt footballsquads.co.uk saadakse kätte info vastava meeskonna mängijate kohta, milleks on:

- mängusärgi number meeskonnas,
- mängija nimi,
- mängija rahvus lühendina (näiteks Saksamaa puhul GER),
- positsioon meeskonnas, nt:
 - G – väravavaht
 - D – kaitsja
 - M – keskväljamängija
 - F – ründaja
- pikkus,
- kaal,
- sünnikuupäev,
- sünnikoht,
- eelmine klubi.

Antud lõputöös kasutatakse vaid osa kättesaadavast infost teksti loomisel (vt ptk-i „Visioon programmi edasi arendamiseks“). Valminud teksti struktuuri määravad ära sünonüümide mallid, mille erinevate vormide valimisel kasutatakse meetodit Random(). Mallid jagunevad järgnevalt:

- meeskondade nimed, kus iga liiga kohta on eraldi mall. Iga rida sisaldab infot ühe meeskonna kohta järgnevalt:
 - märksõna, mille abil muudetakse meeskonna inglisekeelne nimetus eestikeelseks,
 - inglisekeelne nimi, näiteks Bayern Munich,
 - eestikeelne nimi ehk nimetava kääne vorm, näiteks Müncheni Bayern,
 - teksti sees kasutatav meeskonna nime lühend, näiteks Müncheni Bayerni puhul Bayern,
 - omastava käände vormi pikem variant, näiteks Müncheni Bayerni,

- omastava käände vormi lühem variant, näiteks Bayerni,
 - osastava käände vormi pikem variant, näiteks Müncheni Bayernit,
 - osastava käände vormi lühem variant, näiteks Bayernit,
- teksti algus, näiteks „Saksamaa kõrgliigas läksid vastamisi Dortmundi Borussia ja Fraknfurdi Eintracht. Võidu teenis Dortmundi Borussia skooriga 2:0.“, „Saksamaa kõrgliigas Bundesligas kohtusid reedel, 18. augustil omavahel Müncheni Bayern ja Leverkuseni Bayer. Võidu teenis Müncheni Bayern skooriga 3:1.“,
- esimese poolaja esimene värav, näiteks „Avavärava lõi Bayerni poolelt 15. minutil Robert Lewandowski.“, „Mängu läks juhtima Wolfsburg, kui Mario Gomez avas 27. minutil skoori.“,
- jooksvad kirjeldused, näiteks „16. minutil tegi Dortmundi poolelt Mario Götze seisuks 1:0.“, „Järgmise värava eest kandis Dortmundi poolelt hoolt Marco Reus.“,
- teise poolaja esimene värav, näiteks „Teine poolaeg algas Bayerni väravaga, kui 55. minutil lõi värava Arjen Robbe.“, „Poolajapausilt naastes alustas 62. minutil väravaga RB Leipzig, autoriks Timo Werner.“,
- teksti lõpposa, näiteks „Mängu lõppedes jäid tabloole numbrid 3:0.“, „Väravaid enam kohtumise jooksul ei sündinud ja võidu teenis Bayern.“

4.2. Programmi kasutamine ja funktsionaalsused

Programmi kasutamiseks tuleb käivitada NeweratorVersionOne.java fail, olles eelnevalt vajaliku ZIP-faili sisu pakkinud lahti mõnda kindlasse kausta. Ekraanile avaneb aken (vt joonis 3), mille visuaalne pool koosneb viiest komponendist:

- Suvandimenüü „Riik“ – tuleb valida vastav riik, mille põhjal otsitakse välja selle kõrgliiga eesseisvad mängud ja tulemused. Valikute hulgas on viis riiki: Inglismaa, Saksamaa, Hispaania, Itaalia, Prantsusmaa.
- Suvandimenüü „Järgmised mängud“ – siia ilmuvad vastava riigi valimisel selle kõrgliiga eesseisvad mängud, mis pole kasutaja poolt valitavad.
- Suvandimenüü „Tulemused“ – siia ilmuvad vastava riigi valimisel selle kõrgliiga toimunud mängude tulemused, mis on kasutaja poolt valitavad ja vajalikud uudise loomiseks. Eesmärk on anda kasutajale võimalus saada teada mängu toimumise aeg ja näiteks ka veenduda mingi kindla mängu olemasolus, kui seda „Järgmised mängud“ suvandiaknas ei leidu.
- Nupp „Loo uudis“ – selle vajutusel ilmub vastavale tekstiväljale uudise tekst (eelduseks on riigi ja tulemuse valik kasutaja poolt). Samuti käivitub Google Chrome'i veebibrauser, mille ikoon ilmub nähtavale ja kaob paarikümne sekundi pärast. See on tingitud Seleniumi teegi kasutusest.
- Tekstiväli programmiakna keskel – siia ilmub „Loo uudis“ nupu vajutusel genereeritud eestikeelne tekst, mille lõpus on välja toodud hetkeline tabeliseis ja mängu statistika. Fondina on kasutatud „Monospace“ ja kirja suuruseks on 16.

Teksti struktuur on järgnev (vt joonis 5):

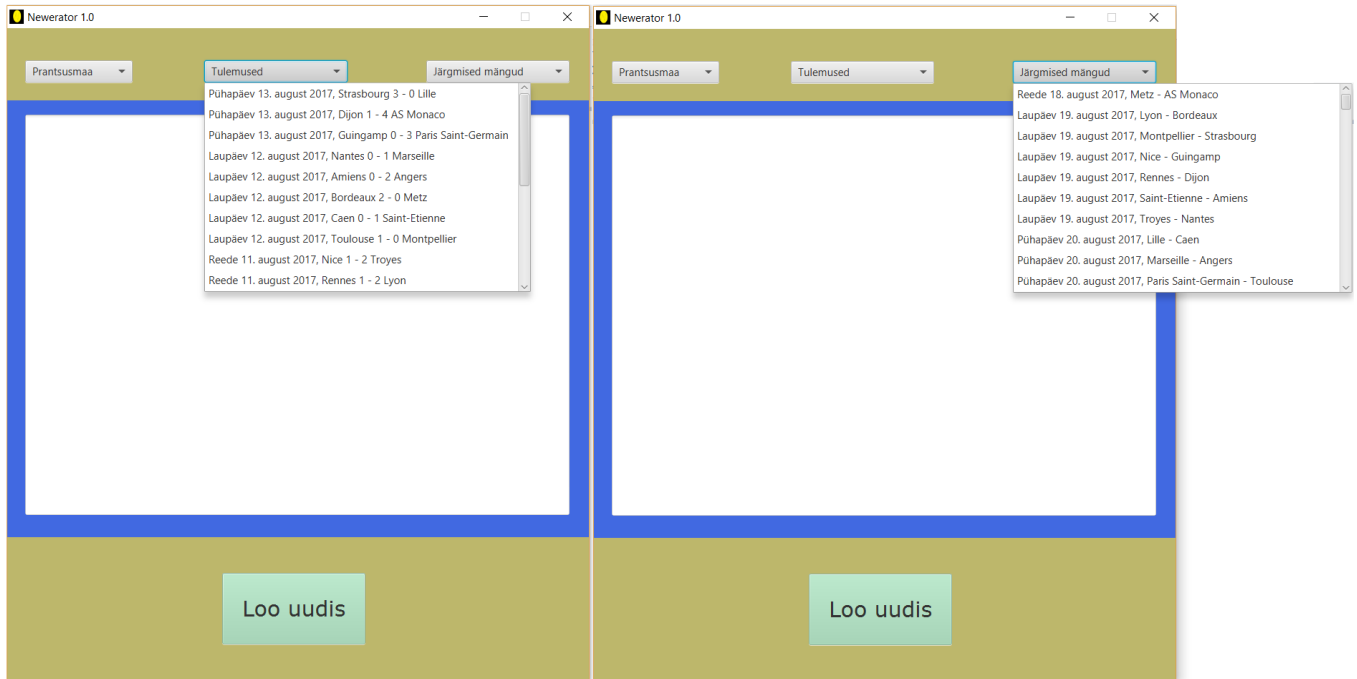
- Esimene lõik on mängu sissejuhatav sõnastus
- Teine lõik annab ülevaate mängu esimesest poolajast (kui selle perioodi jooksul löödi väravaid)
- Kolmas lõik annab ülevaate mängu teisest poolajast (kui selle perioodi jooksul löödi väravaid)
- Peale kolmandat lõiku on kirjas kohtunud meeskonnad kujul „kodumeeskond“ – „külalismeeskond“
- Peale meeskondade nimetusi on mängu statistiline osa
- Viimane osa näitab tabeliseisu, kus:
 - P – punktid

- W – võidud
- D – viigid
- L – kaotused
- VV – löödud ja endale lastud väravate vahe
- Lö – löödud väravate arv
- La – väravate arv, mis on lastud endale lüüa



Joonis 3. Vaade programmi aknast.

Programmiakna suurus on 800x900 pikslit ja selle suurus pole reguleeritav. Kui kasutaja on valinud esimeses suvandiaknas välja riigi, siis tulemuste ja järgmiste mängude alla ilmub vastav sisu (vt joonis 4).



Joonis 4. Prantsusmaa kõrgliiga mängude tulemused (vasakul) ja eesseisvad mängud (paremal).

Näiteks Prantsusmaa kõrgliiga mängu Monaco – Toulouse, mis toimus pühapäeval 4. augustil 2017, tulemusuudise koostamiseks tuleks teha järgmised sammud:

- Käivitada NeweratorVersionOne.java fail
- Avanenud aknas valida suvandimenüü „Riik“ alt „Prantsusmaa“
- Suvandimenüü „Tulemused“ alt valida „Pühapäev 4. märts 2017, AS Monaco – Toulouse“
- Teha hiire vasaku klahvi vajutus nupul „Loo uudis“

Programmi töö kestab umbes 30-40 sekundit sõltuvalt andmehulgast, meeskondade arvust ja toiminud mängude arvust, mis on goal.com-i lehel olemas.



Joonis 5. Programmi poolt loodud uudis 12. augusti 2017 kohta. Vastamisi olid Prantsusmaa kõrgliigas Nantes ja Marseille Olympique. Mäng lõppes tulemusega 1:0 Marseille' poolt vaadatuna.

5. Teadaolevad probleemid ja võimalikud lahendused

Programmi töö ja lõpptulemus sõltub veebist kättesaadavast infost. Loodud on spetsiaalsed meetodid, mis töötlevad HTML-koodi ja suudavad sealt välja otsida vajaliku info.

Uudise koostamine ei õnnestu, kui programmi jaoks infoallikaks oleva veebilehe struktuuri muudetakse. HTML-koodi muutumisel ei pruugi meetod üles leida vajalikke nimesid, tulemusi ja statistikat ning uudise loomine on seeläbi võimatu.

Seleniumi teek oleks sellel juhul lahendus, mis loeb vajalikud andmed sisse viisil, nagu kasutada Ctrl+f kiirklahvi vajutust. Takistuseks võib aga saada selle keerukus teatud andmete kättesaamise puhul. Näiteks kui on veebilehel sama nimega andmeid mitmes kohas, siis tuleks vajalikus meetodis täpsustada, millisel kohal paiknevat nime arvesse võtta. Probleem tekiks samuti veebilehe struktuuri muutumises, kui teatud andmed paigutatakse teistsuguse internetiaadressiga lehele või peidetakse need mõne nupuvajutuse alla.

Programm on täielikult sõltuv internetiühenduse olemasolust, mille puudumisel pole programm võimeline sisse lugema järgmisi mänge, tulemusi ega statistikat. Nendeta kaob võimalus uudist koostada, sest pole lähteandmeid.

Lahendus oleks salvestada kõik andmed kindlasse faili ja kui internetiühendust pole, siis programm kasutaks seda faili uudise loomiseks. Ajaloos on aga leidunud otsuseid, kus mingi mängu tulemus on tühistatud või teatud statistika osa on korrigeeritud. Seega kasutades eraldi faili salvestatud andmeid ei saa alati kindel olla, et uudise sisu vastab reaalsuses tõe.

Lõputöös valminud programmi töös võib esineda vigu mängude puhul, kus väravad on löödud esimesel poolajal ehk esimese 45 minuti raames või pole andmed vajalikul kujul, nagu programmi ülesehitusel eeldatud. Samuti puudub tugi juhtudele, kus esines penaltid¹ [15] ja omaväravaid ning seeläbi genereeritud tekst on puudulik.

¹ Penalti ehk 11-meetri karistuslööök

6. Visioon programmi edasi arendamiseks

Lõputöö autoril on soov arendada valminud programmi edasi. Eesmärgiks on võetud muuta lahendus laiapõhjalisemaks nii valikute kui genereeritava teksti osas.

Antud lõputöö käigus valminud programm kasutab ainult osa olemasolevatest andmetest ja võimalik on uudis luua vaid ühe tulemuse kohta. Edaspidi on eesmärk kasutada kogu infot ja võimaldada uudisesse koondada mitmeid tulemusi, mis muudab tulemusuudise sisukamaks ning mis oleks endiselt kui ühtne tervik. Võimalik on tuua tekstis näiteks välja, mis oli jalgpalluri eelmine koduklubi või kus ta on sündinud.

Praeguse versiooni puhul saab valida viie erineva liiga vahel, kuid see arv võiks laieneda kõigi nende maailmas olevate liigadeni, mille kohta on vähemalt olemas mängude tulemused, väravalööjad, tabeliseisud ja edasised kohtumised käimasoleva hooaja kohta.

Mängu visuaalse poole suhtes puudub programmil võimekus. Kättesaadava statistika põhjal saab teha kahe meeskonna vahel kindlaks aktiivsema poole pealelöökide arvu, pallivaldamise protsendi ja nurgalöökide arvu põhjal. Samuti on võimalik teada saada, kes mängijatest olid väravalööjateks, kuid puudub võimalus kirjeldada, kuidas pall toimetati väravasse või mis täpsemalt eelnes momendile, mil üks meeskondadest lõi värava. Selle võimaluse puudumise puhul näeb lõputöö autor varianti kasutada mõnda võõrkeelset artiklit, uudist või mänguraportit ja tõlkida vastav mängu info eesti keelde. Eelduseks on efektiivne tõlkemehhanism, mis annaks eestikeelse tulemuse võimalikult väheste grammatika ja õigekirja vigadega.

Teine idee on hinnata olukordi tipphetkede video ehk kõige olulisemate sündmuste põhjal, mis mängu jooksul leidsid aset. Eelduseks on luua tarkvara, mis suudab video põhjal eristada kodu- ja külalismseskonna mängijat ning hinnata väravale eelnenud tegevust – milline mängija andis otsustava söödu ehk väravasöödu, kuhu pall löödi värava suhtes vaadates ja kuidas värav löödi (näiteks pealöök, käärlöök, otse õhust ehk vahetult enne palli maandumisel väljakule löök).

„Newerator“ orienteerub vaid ühele spordialale, jalgpallile. Kõigi nende alade suhtes, mille tulemused ja algsemgi statistika on veebis kajastatud, näeb lõputöö autor võimalust laiendada programmi spordialade ringi. Võimalustele seab piiri ala populaarsus: pallimängud ja üldises plaanis meeskondlikud spordialad püüavad aasta lõikes rohkem meedia tähelepanud, kui individuaalsed või viimastel aastatel tekkinud spordialad.

Kokkuvõte

Bakalaureusetöö eesmärk oli koostada programm, mis suudab luua Euroopa jalgpalli viie suurima kõrgliiga – Inglismaa, Saksamaa, Hispaania, Itaalia, Prantsusmaa – raames tulemusuudiseid. Loodud lahendus peaks aitama lihtsustada ajakirjaniku tööd uudiste koostamisel ja säästa tavakasutaja aega erinevatelt veebilehtedelt tulemusuudiseid otsides ja lugedes.

Töö tulemusena valmis Java programmeerimiskeeles kirjutatud programm, mis kogub vajaliku info kolmelt veebilehelt ja kasutades spetsiaalseid malle, luuakse grammatiliselt korrektne eestikeelne tekst.

Valminud programm on algne versioon jalgpalliuudiste loomise osas. Lõputöö autori seisukohalt on võimalik programmi edasi arendada, mis looks veelgi sisukamaid uudiseid ja valikute seas oleks enamus maailma kõrgliigad.

Kasutatud kirjandus

- [1] Sawe , B. E. What Are The Most Popular Sports In The World? Worldatlas, 2017 <http://www.worldatlas.com/articles/what-are-the-most-popular-sports-in-the-world.html> (18.05.2017)
- [2] Laws of the Game 2016/2017. FIFA http://www.fifa.com/mm/Document/FootballDevelopment/Refereeing/02/79/92/44/Laws.of.the.Game.2016.2017_Neutral.pdf
- [3] History of Football – The Origins, fifa.com <https://www.fifa.com/about-fifa/who-we-are/the-game/> (18.05.2017)
- [4] Jalgpalliuudiste portaal soccernet.ee <http://soccernet.ee/>
- [5] Kuldkepp, A. Messi tõi Barcelonale kolm punkti ja kullavõimalus jäi alles. Postimees. 16. aprill 2017. <http://sport.postimees.ee/4081211/messi-toi-barcelonale-kolm-punkti-ja-kullavoimalus-jai-alles> (18.05.2017)
- [6] Nõmm, Di Maria duubel jättis Prantsusmaal tiitliheitluse lahtiseks. Eesti Rahvusringhääling. 15. aprill 2017. <http://sport.err.ee/590017/di-maria-duubel-jattis-prantsusmaal-tiitliheitluse-lahtiseks> (18.05.2017)
- [7] Teegi Selenium koduleht <http://www.seleniumhq.org/> (14.08.2017)
- [8] Jalgpalliuudiste portaal goal.com <http://www.goal.com>
- [9] Uudisteportaal skysports.com <http://www.skysports.com/football>
- [10] Jalgpallialane andmebaas footballsquads.co.uk <http://www.footballsquads.co.uk> (14.08.2017)
- [11] N. D. Allen, J. R. Templon, P. S. McNally, L. Birnbaum, K. Hammond, StatsMonkey: A Data-Driven Sports Narrative Writer – AAAI Publications, 2010 AAAI Fall Symposium Series, Computational Models of Narrative, Evanston, IL, 2010.
- [12] Narrative Science, hired.com <https://hired.com/companies/narrative-science> (18.05.2017)
- [13] Program Creates Computer-Generated Sports Stories, npr.org <http://www.npr.org/templates/story/story.php?storyId=122424166> (18.05.2017)

[14] M. Molina, Simulating Data Journalism to Communicate Hydrological Information from Sensor Networks. Advances in Artificial Intelligence – IBERAMIA 2012. Lecture Notes in Computer Science, vol 7637. Springer, Berlin, Heidelberg, 2012, lk 722-731

[15] Laws of the Game 2015/2016. FIFA
https://www.fifa.com/mm/Document/FootballDevelopment/Refereeing/02/36/01/11/LawsofthegamewebEN_Neutral.pdf, lk 45-48

Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, **Martin Reinsalu**,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

Jalgpalliuudiste koostaja „Newerator“ algversioon,

mille juhendaja on **Krista Liin**,

1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **14.08.2017**