Professional Paper | Received: 05-10-2016 | Accepted: 17-05-2017

Semi-Automatic Story Generation for a Geographic Server

Rizwan MEHMOOD and Hermann MAURER

Graz University of Technology, Rechbauerstraße 12, Graz, Austria
rmehmood@iicm.edu, hmaurer@iicm.edu

Abstract. Most existing servers providing geographic data tend to offer various numeric data. We started to work on a new type of geographic server, motivated by four major issues: (i) How to handle figures when different databases present different values; (ii) How to build up sizeable collections of pictures with detailed descriptions; (iii) How to update rapidly changing information, such as personnel holding important functions, and (iv) how to describe countries not just by using trivial facts, but stories typical of the country involved. We have discussed and partially resolved issues (i) and (ii) in previous papers; we have decided to deal with (iii), regional updates, by tying in an international consortium whose members would either help themselves or find individuals to do so. It is issue (iv), how to generate non-trivial stories typical of a country, that we decided to tackle both manually (the consortium has by now generated around 200 stories), and by developing techniques for semi-automatic story generation, which is the topic of this paper. The basic idea was first to define sets of reasonably reliable servers that may differ from region to region, to extract "interesting facts" from the servers, and combine them in a raw version of a report that would require some manual cleaning-up (hence: semi-automatic). It may sound difficult to extract "interesting facts" from Web pages, but it is quite possible to define heuristics to do so, never exceeding the few lines allowed for quotation purposes. One very simple rule we adopted was this: 'Look for sentences with superlatives!' If a sentence contains words like "biggest", "highest", "most impressive" etc. it is likely to contain an interesting fact. With a little imagination, we have been able to establish a set of such rules. We will show that the stories can be completely different. For some countries, historical facts may dominate; for others, the beauty of landscapes; for others, cultural or economic achievements, and for yet others, unusual facts concerning Nobel Prize winners, food, entertainment, sports, other activities, national symbols, special laws, and so on. The results can be checked on by clicking on any country in the category "Special Information" under "Surprising Facts". All examples shown in this paper were chosen fairly arbitrarily from over 190 examples, to show that the system is indeed working. There are two points to mention here: (a) it is a work in progress, yet has reached a very useable size; (b) the basic ideas can be applied to any area. The choice of geography was due to the wealth of data and interest in this area, but if our algorithms overlook some important facts, this is less critical than applied to types of medical treatment, etc.

Keywords: story generation, geographic server

1 Introduction

Geographic facts influence world history, society and human development. Geographic facts offer students, teachers, researchers, and the general public tools for better understanding our world. The Web has enormous amounts of data from which one can easily extract information of interest. Typically, online encyclopaedias such as Wikipedia provide information about all the countries of world. Their objective is usually to provide as much information about a particular country as can be collected.

This is what sometimes is wanted. In other cases, a short overview, or an emphasis on a particular topic might be more desirable, or just some highlights typical of the country. Note that any lengthy information in Wikipedia is preceded by a short summary, but this is not enough to cater for the very different interests of users. As long as we cannot specify something like "I want an n character long exposition on country x with a

Stručni rad | Primljeno: 05-10-2016 | Prihvaćeno: 17-05-2017

Poluautomatsko stvaranje priča/sadržaja za geografski poslužitelj

Rizwan MEHMOOD i Hermann MAURER

Tehničko sveučilište u Grazu, Rechbauerstraße 12, Graz, Austrija
rmehmood@iicm.edu, hmaurer@iicm.edu

Sažetak. Većina postojećih poslužitelja koji nude geografske podatke sadrže brojčane podatke o različitim aspektima. Na rad na novom tipu geografskog poslužitelja potaknula su nas četiri glavna problema: (i) Kako se služiti brojkama kad različite baze podataka predstavljaju različite vrijednosti; (ii) Kako izgraditi velike zbirke slika s detaljnim opisima; (iii) Kako ažurirati informacije koje se brzo mijenjaju kao što su osobe na nekim važnim funkcijama te (iv) Kako opisati zemlje ne samo trivijalnim činjenicama, već tipičnim pričama za te zemlje. U prethodnim smo radovima raspravili i djelomično riješili probleme (i) i (ii). Odlučili smo riješiti (iii), regionalna ažuriranja, povezivanjem s međunarodnim konzorcijem čiji će članovi pomoći ili naći pojedince koji će to učiniti. Problemu (iv), kako stvoriti netrivijalne priče tipične za neku zemlju, pristupili smo ručno (konzorcij je do sada stvorio oko 200 priča) te razvijanjem tehnika za poluautomatsko stvaranje priča, što je tema ovoga rada. Osnovna je ideja bila prvo odrediti skupove pouzdanih poslužitelja koji se mogu razlikovati od regije do regije, izvući „zanimljive činjenice“ iz njih i spojiti ih u sirovu verziju izvještaja koja će se obraditi ručno (zato ga nazivamo poluautomatskim). Izdvajanje „zanimljivih činjenica“ s internetskih stranica može zvučati teško, no itekako je moguće odrediti heuristike koji će to učiniti, što nikad ne prelazi nekoliko redova za svrhu citiranja. Spomenimo kao primjer jedno vrlo jednostavno pravilo: Traži rečenice sa superlativima! Ako rečenica sadrži riječ kao što je „najveći“, „najviši“, „najimpresivniji“ i sl., ona vjerojatno sadrži neku zanimljivu činjenicu. S pomoću mašte uspjeli smo odrediti skup takvih pravila. Pokazat ćemo da priče mogu biti potpuno različite: u nekim zemljama dominiraju povijesne činjenice, u drugima ljepota krajolika, u trećima kulturna i ekonomska postignuća, u nekima neobične činjenice koje se odnose na dobitnike Nobelove nagrade, hranu, sport, druge aktivnosti, državne simbole, posebne zakone i sl. Dobiveni rezultati mogu se provjeriti traženjem bilo koje zemlje u kategoriji „Posebne informacije“ (Special Information) i „Iznenađujuće činjenice“ (Surprising Facts). Svi primjeri opisani u ovom radu uzeti su arbitrarno iz skupine od 190 primjera kako bi se prikazalo kako sustav radi. Važno je spomenuti još dvije stvari: (a) riječ je o radnoj verziji koja je već prilično upotrebljiva; (b) osnovne ideje mogu se primijeniti na bilo koje područje. Geografija je izabrana s obzirom na velik broj podataka i interesa za to područje. Ako su naši algoritmi previdjeli neku važu činjenicu, to je manje važno nego da smo primijenili metode na vrste liječenja ili slično tome.

Ključne riječi: stvaranje priča, geografski poslužitelj

1. Uvod

Geografske činjenice utječu na svjetsku povijest, društvo i ljudski razvoj. One omogućuju učenicima, učiteljima, istraživačima i javnosti alate kojima mogu bolje razumjeti naš svijet. Internet sadrži ogromnu količinu podataka iz kojih pojedinac može izdvojiti informacije koje ga zanimaju. Enciklopedije na internetu (kao što je Wikipedia) nude informacije o svim zemljama svijeta. Cilj im je pružiti što više informacija o pojedinoj zemlji.

U nekim je slučajevima pak poželjniji kratak pregled, naglasak na nekoj temi ili činjenicama tipičnima za određenu zemlju. Premda svakom dugačkom nizu informacija na Wikipediji prethodi sažetak od nekoliko redova, to nije dovoljno za mnogobrojne različite interese korisnika. Dok god ne možemo reći nešto kao *Želim izlaganje o zemlji x u n znakova s velikim naglaskom na temu y*, potrebno je na druge načine zadovoljiti različite ukuse ili interese, ovisno o potrebi. Spomenut ćemo neke važne činjenice i druge pokušaje, a nakon toga objasniti svoj pristup.

distinct emphasis on topic y" other avenues have to be used to satisfy various tastes or interests. We will explain our approach after mentioning some important facts and other attempts.

Wikipedia is a remarkable resource as a corpus for knowledge extraction (Nakayama et al. 2008). Even if not perfect, its quality has often been examined with satisfactory results. There have been some efforts in the past to extract information from Wikipedia (Medelyan et al. 2004). Recent years have seen the use of computers in data mining for extracting hidden knowledge; see (Yeung, Jatwot 2011). The authors in (Yeung, Jatwot 2011) have tried to highlight important years for different nations. They performed their experiments on the Google news archive. They identified frequently mentioned years for different countries, like 1974, 1976, and 1978 for Argentina. The top words were "team", "first", "cup", and "Maradona "(a footballer); the reasons clearly being the success stories of Argentina's soccer team in those years.

Let us now turn to the approach that we are taking. The web pages for a particular country in our geographic server cover information on different aspects of geography. Note that by adding a special section on maps, culture, pictures and stories, we allow users to pick the parts that are of interest to them. We discussed this in the paper (Mehmood et al. 2016).

The current paper focuses on stories that are typical of a country, are semi-automatically generated, and can be found in the category "Special Information" for each country as "Surprising Facts" (URL1). However, to get a substantial quantity of facts, we used many databases, like Factbook, DBpedia, the World Health Organization, World Trade Association and other data sources mentioned in the references. We also used textual information on countries from Wikipedia. We explicitly highlight (marked in italics) and properly referenced items taken from Wikipedia or other sources that require reference to some licence.

To give an idea how we try to discover "interesting facts", let us present a few examples. As mentioned before, sentences with superlatives usually contain interesting information; sentences with a date (a year) usually indicate something special happened at that point in time, which can be further verified by finding the same date on other web pages, or if the date is associated with a special word like "revolution", "war", "liberation", "freedom", "invention", etc. For all the UN countries, our database has entries covering more than 100 properties. When a country is very high/ low in one of the properties, this calls for attention. When looking at the very reliable Nobel Prize site (URL13), it is clear that some countries are prominent in one way or another. The same holds for sites listing artists or Olympic medallists. Other interesting facts include the position of a country, its climate, variation in languages, or whether it has enclaves or exclaves. Uzbekistan has a number of enclaves in Kyrgyzstan, Kaliningrad is part of Russia, but has no border with Russia, while it has them with Lithuania, Belarus and Poland. Even Alaska and Hawaii, important, yet far away from the US mainland, are interesting cases.

Maps are useful for grasping geographic facts. Maps of neighbouring countries are provided to promote the easy understanding of facts related to boundaries and seashores. The maps in stories are fully interactive and users can hover over a country to see the population of selected and neighbouring countries. Freely available mapping libraries such as Openlayers (URL16) and Leaflet (URL17) are empowering developers to use the Web as a medium for displaying maps. Adding population statistics to maps helps users to comprehend a country within a geographic context. We have used Leaflet to display maps. The tiles are taken from the tile provider, CartoDB (URL18).

Another example of a typical surprising fact might be that Austria not only imports and exports to and from many European countries, but also imports surprisingly large quantities of products from unexpected places like Kazakhstan. Similarly, some Asian countries, like Sri Lanka and Israel, export more than one would expect to Belgium. The exports commodities of Sri Lanka are textiles, clothes, tea, spices, rubber products, precious stones, etc. The African country Burkina Faso exports to Belgium; the exports commodities are gold, cotton etc. An interesting fact is the number of patents registered by a country. For instance, the United States has the largest number of patents (157,496). Japan is in second position (125,880), see (URL3).

It is interesting and even entertaining to discover that a country got a Nobel Prize for a particular area unexpectedly early or late: e.g. the German-born scientist Wilhelm Conrad got the Nobel Prize for physics in the first year (1901) Nobel Prizes were awarded. France has the largest number of Nobel Prize winners (counting those was born in France) in literature (11). Marie Curie, who was born in Poland, was the first woman to get a Nobel Prize for physics. She was awarded it in 1903. We generate stories by extracting information from different databases and taking only those that match specific constraints. Our overall objective is to entertain or surprise people by summarizing some of the most important and unique facts.

This paper presents an approach to acquire exciting knowledge from various data sources about countries,

Wikipedia je izvanredna zbirka znanja (Nakayama i dr. 2008). Iako nije savršena, istraživanja često pokazuju da je njezina kvaliteta zadovoljavajuća. Već postoje neki pokušaji izdvajanja informacija iz Wikipedije (Medelyan i dr. 2004). U posljednje se doba računala upotrebljavaju za izdvajanje skrivenoga znanja (Yeung, Jatwot 2011). Yeung i Jatwot (2011) pokušali su istaknuti važne godine za različite zemlje. Svoje su istraživanje proveli na Googleovoj arhivi vijesti. Utvrdili su godine koje se često spominju za različite zemlje, npr. 1974., 1976. i 1978. za Argentinu, a glavne riječi bile su "tim, prvi, kup, Maradona (igrač nogometa)",a razlog tome su očito priče uspjeha argentinskog nogometnog tima u tim godinama.

Okrenimo se sad pristupu koji smo primijenili. Web stranice za određenu zemlju u našem geografskom poslužitelju pokrivaju informacije o različitim aspektima geografije. Dodavanje posebnog dijela o kartama, kulturi, slikama i pričama omogućuje korisnicima da izaberu ono što ih zanima (Mehmoof i dr. 2016).

Kao što smo spomenuli, ovaj se rad usmjerava na priče koje su tipične za određenu zemlju, stvorene su poluautomatski i mogu se pronaći u kategoriji „Posebne informacije" (Special Information) u svakoj zemlji kao „Iznenađujuće činjenice" (Surprising Facts) (URL1). Međutim, kako bismo dobili veliku količinu činjenica, upotrijebili smo mnoge baze podataka, kao što su Factbook, DBpedia, World Health Organisation, World Trade Association i druge izvore podataka koji su navedeni u popisu literature. Također smo upotrijebili tekstualne informacije o zemljama iz Wikipedije. Posebno naglašavamo (kurzivom) i citiramo tekstove koje smo preuzeli s Wikipedije ili iz drugog izvora koji zahtijeva citiranje.

Kako bismo ilustrirali kako smo došli do „zanimljivih činjenica", opisat ćemo nekoliko primjera. Kao što smo spomenuli, rečenice koje sadrže superlativ obično sadrže i zanimljivu informaciju; rečenice koje sadrže datum (godinu) obično ukazuju da se tada dogodilo nešto važno. To je moguće provjeriti pronalaženjem istog datuma na drugim web stranicama ili ako se datum nalazi kraj posebne riječi kao što je „revolucija", „rat", „oslobođenje", „sloboda", „izum" i dr. Naša baza podataka sadrži više od 100 svojstava svih članica UN-a. Ako je neka zemlja vrlo visoko ili nisko na nekom od tih svojstava, to privlači pažnju. Ako pogledamo vrlo pouzdanu stranicu o dobitnicima Nobelove nagrade (URL13), očito je da se neke zemlje izdvajaju na jedan ili drugi način. Isto vrijedi za stranice s popisima umjetnika ili dobitnika olimpijske medalje. Zanimljive su i činjenice o položaju pojedine zemlje, njezinoj klimi, jezicima te gdje se nalaze njezine enklave ili eksklave. Tako Uzbekistan ima velik broj enklava u Kirgistanu, Kalinjingrad je dio Rusije, iako nema granicu s Rusijom, ali graniči s Litvom, Bjelorusijom i Poljskom. Zanimljivi su i Aljaska i Havaji, važni dijelovi Sjedinjenih Američkih Država, udaljeni od kontinentalnog dijela SAD-a.

Karte su važne za shvaćanje geografskih činjenica. Za lakše razumijevanje činjenica koje se odnose na granice i obale dana je karta susjednih zemalja. Karte u pričama potpuno su interaktivne i korisnici mogu vidjeti naseljenost izabranih zemalja i njihovih susjednih zemalja. Besplatne biblioteke za kartiranje kao što su Openlayers (URL16) i Leaflet (URL17) omogućuju programerima upotrebu weba kao medija za prikazivanje karata. Dodavanje statističkih podataka o stanovništvu kartama omogućuje korisnicima shvaćanje određene zemlje u geografskom kontekstu. Za prikazivanje karata upotrijebili smo Leaflet. Pločice (tiles) smo preuzeli iz CartoDB-a (URL18).

Drugi primjer tipične iznenađujuće činjenice je taj da Austrija ne samo da uvozi iz mnogih europskih zemalja i da u njih izvozi, već uvozi iznenađujuće velike količine proizvoda iz neočekivanih zemalja kao što je Kazahstan. Slično tome, azijske zemlje poput Šri Lanke i Izraela izvoze više nego što bi se očekivalo u Belgiju. Šri Lanka izvozi tekstil, odjeću, čaj, začine, gumu, dragocjeno kamenje, itd. Afrička zemlja Burkina Faso izvozi u Belgiju zlato, pamuk, itd. Također je zanimljiva činjenica o broju patenata registriranih u pojedinoj zemlji. Na primjer, SAD ima najveći broj patenata (157 496), dok je Japan drugi s 125 880 (URL3).

Zanimljivo je i zabavno saznati da je neka zemlja dobila Nobelovu nagradu u nekom području neočekivano rano ili neočekivano kasno: npr. Wilhelm Conrad, rođen u Njemačkoj, dobio je Nobelovu nagradu za fiziku 1901. kad su se počele dodjeljivati Nobelove nagrade. Francuska ima najveći broj Nobelovih nagrada za književnost (11). Marie Curie bila je prva žena rođena u Poljskoj koja je dobila Nobelovu nagradu za fiziku 1903. godine. Stvaramo priče izdvajanjem informacija iz različitih baza podataka, pri čemu ih uzimamo samo ako zadovoljavaju određene uvjete. Naš je opći cilj zabaviti ili iznenaditi ljude sažimanjem najvažnijih i najneobičnijih činjenica.

U ovom je radu primijenjen pristup izdvajanja znanja iz različitih izvora o zemljama, što ćemo podrobnije opisati u 3. poglavlju. Ostatak rada organiziran je na sljedeći način: u 2. poglavlju objasnit ćemo arhitekturu sustava, a u 3. staviti žarište na heuristike. Završit ćemo zaključkom i kratkim popisom reprezentativnih izvora.

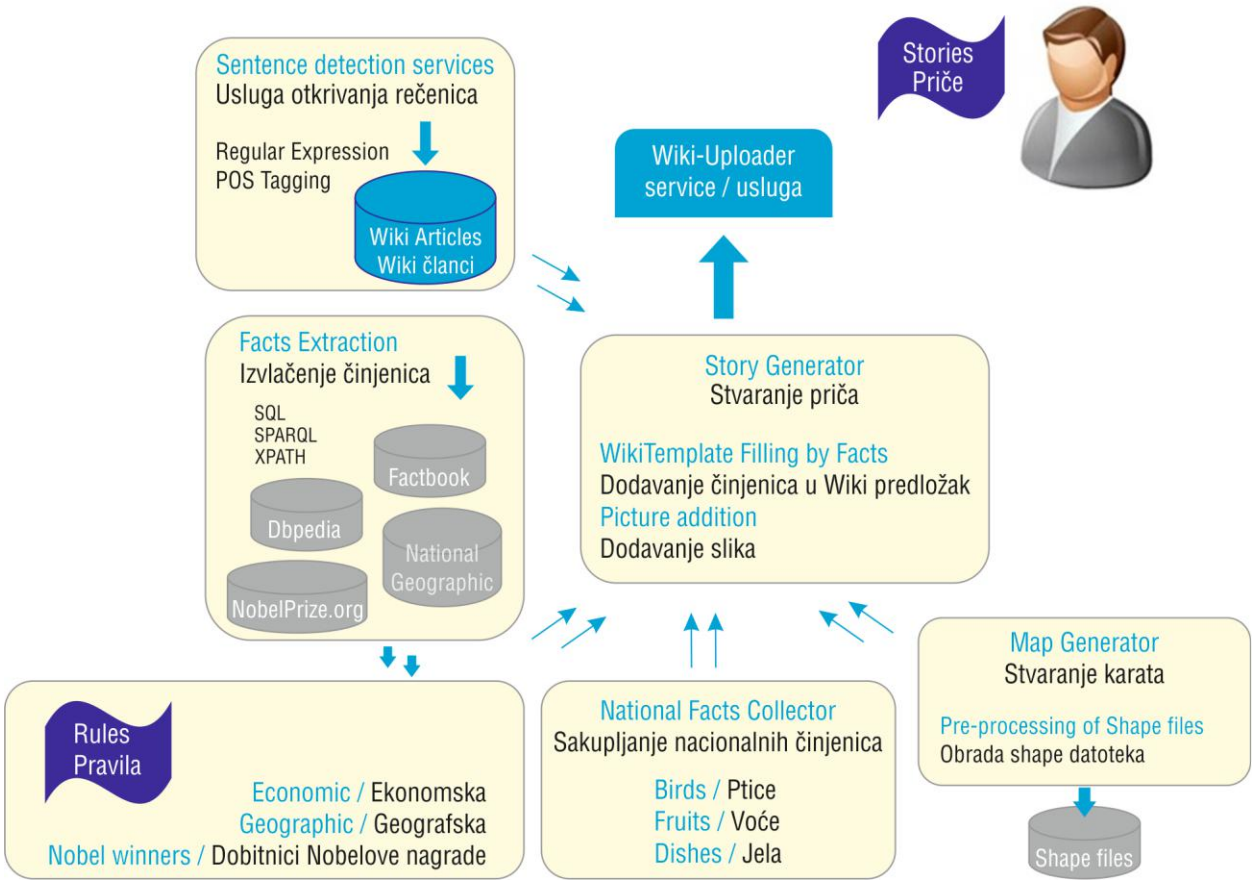

Fig. 1 Overall architecture of the system (We have shown here only a few of the data sources used as examples)
Slika 1. Opća arhitektura sustava (samo nekoliko izvora podataka navedeno je kao primjer)

as we will explain in Section 3. The rest of the paper is organized as follows. We will explain the architecture of the system in Section 2. We will narrow our focus in Section 3 and describe our heuristics in corresponding subsections in detail. We end with a conclusion and a short list of representative references.

2 Architecture of the System

Figure 1 shows the overall architecture of story generation. We start the story of each country using sentences extracted from Wikipedia articles. For the extraction of geographic and economic facts, we use the CIA World Factbook (URL19). Further, we use facts from DBpedia, National Geographic (URL20) (travel-guide page), and the Nobel Prize website (URL13). The map generator module shown in Figure 1 is responsible for the dynamic maps displayed along with (hopefully) amusing descriptions. We have downloaded the shape file (.shp) of the countries from Natural Earth (URL21). To select only the neighbouring countries of a particular country used "Select Features by Free hand" (a special QGIS feature that allows the selection of particular polygons). We have extracted these specific polygons and used them in the corresponding maps of countries. The map of France is shown in Figure 2.

The last step was the aggregation of different facts in a template designed for stories. Once we had stories in a final form, we augmented them with suitable pictures and videos. This process is currently being done manually, requiring some effort. We hope to cut this down by using a more automatic approach, as we did when merging picture databases, see (Mahmood, Maurer 2015).

3 Heuristics for Story Generation

As explained, we are trying to present for each country interesting pieces of information beyond maps and information that can be easily deduced from the numeric data in our databases.

There are three completely different types of information that we believe we should provide:

(i) Information that is intuitively immediately associated with a country

2. Arhitektura sustava

Slika 1 prikazuje pregled arhitekture stvaranja priča. Priču svake zemlje počinjemo rečenicama izvučenim iz članaka na Wikipediji. Za izvlačenje geografskih i ekonomskih činjenica primijenili smo CIA World Factbook (URL19). Nadalje, upotrebljavamo činjenice iz DBpedije, National Geographica (URL20) (stranica vodiča za turiste) i stranice o dobitnicima Nobelove nagrade (URL13). Modul stvaranja karata, prikazan na slici 1, odgovoran je za dinamičke karte koje se prikazuju s, nadajmo se, zabavnim opisima. Datoteku oblika zemalja (.shp) preuzeli smo s Natural Eartha (URL21). Kako bismo označili samo susjedne zemlje, upotrijebili smo funkciju *Select Features by Free hand* (posebnog svojstva QGIS-a koje omogućuje označavanje određenih poligona). Izdvojili smo te poligone i upotrijebili ih na odgovarajućim kartama zemalja. Karta Francuske prikazana je na slici 2.

Posljednji je korak bio skupiti različite činjenice u predložak za priče. Kad smo dobili priče u konačnom obliku, poboljšali smo ih odgovarajućim slikama i videoisječcima. Taj se postupak trenutno radi ručno, što iziskuje određen napor. Nadamo se da ćemo taj postupak olakšati primjenom pristupa koji je u većoj mjeri automatiziran, kao što smo to učinili sa spajanjem baza podataka o slikama (Mehmood, Maurer 2015).

3. Heuristike za stvaranje priča

Svaku zemlju pokušavamo predstaviti zanimljivim informacijama uz karte i informacije o kojima se može lako zaključiti iz brojčanih podataka u našim bazama podataka.

Smatramo da trebamo pružiti tri potpuno različita tipa informacija:

(i) informacije koje se intuitivno odmah povezuju sa zemljom
(ii) informacije koje su službeno povezane sa zemljom, a možda nisu općepoznate
(iii) informacije koje su tipične za zemlju, ali slabo poznate.

Prije nego što objasnimo kako možemo doći do takvih informacija, navest ćemo nekoliko primjera za tipove informacija (i) – (iii). Što se tiče tipa informacija (i), smatramo da osobu s određenom zemljom povezuju gotovo sve ove riječi (namjerno posložene nasumično): koala, ptica kivi, noj, most Golden Gate, Slapovi Niagare, Eiffelov toranj, kosi toranj u Pisi, piramide, Uluru, Big Ben, panda, Kineski zid, Fuji, Kilimandžaro, šampanjac, Loch Ness, izumitelj parnog stroja, Hirošima, Mt. Everest, slatkokisela juha, bečki odrezak, grizli, Mozart, Jukatan, filozof Kant ili psihijatar Sigmund Freund (i stotinu drugih pojmova). Isto vrijedi za neke slike. Tako su vrlo prepoznatljive slike Stonehengea, para koji pleše tango ili pak Manhattana.

Što se tiče (ii), zemlje često proglašavaju nacionalne simbole kao što su cvijet, voće, životinja, ptica ili hrana, no često te činjenice ne znaju niti stanovnici tih zemalja.

Mnogi Austrijanci ne znaju da je lasta nacionalna ptica njihove zemlje, Armenci da je marelica nacionalno voće njihove zemlje ili Pakistanci da je jarebica nacionalna ptica njihove zemlje. No potrebno ih je navesti kad se opisuje neka zemlja.

Posebno je zanimljiva skupina (iii): koliko ljudi zna da neobična biljka welwitschia raste samo u Namibiji, da automobili u hladnim dijelovima Kanade trebaju imati ugrađen grijač ulja kako bi zimi mogli voziti, da pod pustinjom u Zapadnoj Australiji postoji more, da je u toj zemlji posebno stablo upotrebljavano kao zatvor (URL4), da je zmajevo stablo u Tenerifima poznato po ispuštanju krvi (crvene smole), da je tek 1963. u Butanu otvorena prva cesta za vozila ili da u Europi dugo nije bilo dozvoljeno pušiti na otvorenom. Opisat ćemo kako smo do navedenih informacija na internetu došli.

3.1. Nacionalne posebnosti/činjenice/znakovi

Pogledajmo neke nacionalne činjenice. Runolist je nacionalni cvijet Austrije. Ruža je nacionalni cvijet u SAD-u, Ujedinjenom Kraljevstvu, itd. (URL5). Kriket je nacionalni sport u američkim državama kao što su Antigua i Barbuda, Barbados i Bermudski otoci. Nogomet je nacionalni sport u europskim zemljama poput Mađarske i Poljske. Ples je umjetnost koja predstavlja nacionalnu kulturu. Na primjer, kolo je nacionalni ples u Hrvatskoj, a Landler je narodni ples popularan u Austriji. Zemlje predstavljaju i ptice i druge životinje. Indija je poznata po različitim vodenim životinjama, pticama, gmazovima, itd. Kraljevska kobra smatra se nacionalnim gmazom u Indiji. Indijski je slon baštinska životinja. Bjelorusija je poznata po bijelim rodama.

3.2. Turističke atrakcije

Turistima su često zanimljive pećine. Pećina svijetlećih crva Waikato postala je poznata zbog pogrešnog razumijevanja. Drugi primjer su velike ledene pećine poput onih u Salzburgu (Austrija). Toj kategoriji pripadaju i posebni mostovi. Na primjer, most Danyang-Kunshan u Kini je najdulji most na svijetu (URL6). Muzej Taxila poznata je turistička atrakcija u Pakistanu. Louvre je najposjećeniji muzej na svijetu (više od 12 milijuna posjeta godišnje), a prema površini je treći na svijetu.

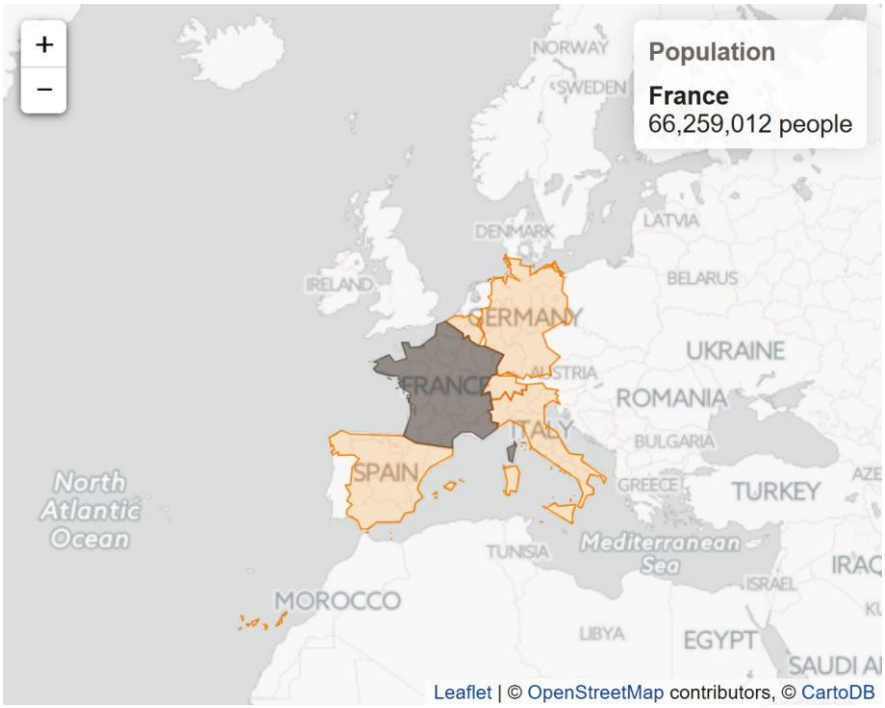

Fig. 2 Map showing neighbouring countries of France. Population is shown when in the system by hovering with the mouse over a particular area

Slika 2. Karta prikazuje zemlje susjedne Francuskoj. Pomicanjem miša preko određene zemlje prikazuje se njezina naseljenost

(ii) Information that is officially associated with a country, yet may not be universally known
(iii) Information that is typical of a country, yet is little known.

Let us first present some examples of (i) – (iii) before discussing how we can obtain the pertinent information. Concerning (i), we believe that almost all of the words (on purpose arranged very unsystematically) like koala bear, kiwi bird, ostrich, Golden Gate Bridge, Niagara Falls, Eiffel Tower, Leaning Tower of Pisa, pyramids, Ayers Rock, Big Ben, panda, Great Wall, Mount Fuji, Kilimanjaro, Champagne, Loch Ness, inventor of steam engine, Hiroshima, Mt. Everest, sweet and sour soup, Wiener Schnitzel, grizzly bear, Mozart, Yucatan, the philosopher Kant or psychiatrist Sigmund Freund (and hundreds more words) are immediately associated with a country or area of the world. The same is true of some pictures. Pictures of Stonehenge, a couple dancing the tango, or the skyline of Manhattan, etc., will be recognized and associated with some area of the world by most people immediately.

Concerning (ii), countries often declare items as national symbols, such as flowers, fruit, animals, birds, or food items, yet even the inhabitants of the country are often not aware of such facts.

Many Austrians would not know that the swallow is the national bird of Austria, or Armenians that the apricot is the national fruit of Armenia, or Pakistanis that the Chukar Partridge is the national bird of Pakistan. Yet when describing a country, such items should be listed.

Group (iii) is particularly interesting: How many people know that a strange plant called Welwitschia grows only in Namibia, that cars in cold areas of Canada need a plug-in (an oil heater) to keep working in winter, that there is a veritable sea of (unfortunately brackish) water under the desert of Western Australia, that a certain tree "Boab Prison Tree" in the same country was used as an overnight prison (URL4), that the dragon tree of Tenerife is famous for releasing "blood" (red sap) when cut, that the first road used by wheeled vehicles in Bhutan was opened only in 1963, or that smoking outside buildings was not allowed in Europe for a long time. We are now going to describe how we try to dig out such information from what is available on the web, sometimes supported by the community.

3.1 National Goods/Facts/Symbols

Let us look at some national facts. The edelweiss is the national flower of Austria. The rose is considered the national flower in the United States and United Kingdom etc. (URL5). American countries such as Antigua and Barbuda, Barbados, and Bermuda are fond of cricket, which is considered the national sport in these countries. Soccer is considered the national sport in European countries such as Hungary and Poland. Dance is a type of art which represents the culture of a nation. For instance, the *kolo* is considered the national dance in Croatia. The *landler* is a folk dance which was very popular in Austria, The Viennese waltz is a genre of ballroom dance popular in Austria. Birds and animals also

Fig. 3 Map showing Afghanistan is a landlocked country

Slika 3. Karta prikazuje da Afganistan nema izlaz na more

3.3. Filtriranje činjenica s pomoću pravila

S pomoću pravila izabrali smo ekstremne vrijednosti (superlative). Pakistan izvozi u različite zemlje, ali jedinstvenost proizlazi iz primjene pravila max (izvoz). Prvo smo izdvojili postotak izvoza iz skupa Factbook dataset (URL7) te potom utvrdili zemlju najvećeg izvoza: Pakistan najviše izvozi u SAD, a najviše uvozi iz Kine.

3.4. Geografske priče

Slijede neke zanimljive geografske činjenice tipične za određenu zemlju:

- okružena kopnom
- najveća granica s određenim susjedom
- najdulja rijeka
- bogatstvo određenom sirovinom (dijamanti u Južnoj Africi, nafta u nekim zemljama, ...).

Zanimljiva je činjenica ima li zemlja izlaz na more ili ne, npr. Švicarska i Austrija ga nemaju. Tu je činjenicu moguće provjeriti pregledom karata koje se nalaze u pričama. Afganistan nema izlaz na more, što je prikazano na slici 3.

Premda većina zemalja graniči s drugim zemljama, zanimljivo je vidjeti koja zemlja ima najdulju ili najkraću granicu s odabranom zemljom. Afganistan ima najdulju granicu s Pakistanom, otprilike 2670 km. Najkraća granica Afganistana je ona s Kinom (91 km), što je prikazano na slici 3.

3.5. Određivanje jedinstvenih činjenica iz članaka na Wikipediji s pomoću neurolingvističkog programiranja (NLP-a)

Osvrnut ćemo se na neke ideje o tome kako pronaći relevantne i zanimljive činjenice uz pomoć obrade teksta, što smo upotrijebili u uvodnom dijelu ove priče. Izvukli smo dva tipa rečenica:

1. rečenice sa superlativima/pridjevima
2. rečenice s godinama (za izdvajanje povijesnih događaja).

3.5.1. Utvrđivanje zanimljivih rečenica

Slijedi nekoliko zanimljivih rečenica o zemljama iz članaka na Wikipediji.

Honduras ima najvišu stopu ubojstva na svijetu.

Kuba je godine 2015. postala prva zemlja na svijetu koja je iskorijenila prijenos HIV-a i sifilisa s majke na dijete, što je prekretnica koju je Svjetska zdravstvena organizacija nazvala „jednim od najvećih mogućih dostignuća u javnom zdravstvu“.

Grenada je poznata i kao „Otok začina“ zbog proizvodnje muškatnog oraščića i jedna je od najvećih izvoznica tog začina.

Gana je jedan od najvećih proizvođača zlata i dijamanata na svijetu i predviđa se da će u 2015. postati najveći proizvođač kakaa na svijetu.

Papua Nova Gvineja jedna je od najmanje istraženih zemalja na svijetu u kulturološkom i geografskom

represent countries. India is known for different animals including aquatic birds, reptiles etc. The king cobra is considered the national reptile in India. The Indian elephant is a heritage animal. Belarus is famous for its white storks.

3.2 Tourist Attractions

Concerning tourist destinations, caves are often of interest. The Waikato glow-worm caves became famous because of a curious misunderstanding. Other examples are large ice caves such as those in Salzburg (Austria). Items such as special bridges also belong in this category. Danyang–Kunshan Grand Bridge is the world longest bridge in China (URL6). Taxila Museum is one of the famous tourist attractions in Pakistan. France has the largest art museum (the Louvre Museum) in the world, measured by yearly attendance (over 12 million per year); it is the third largest by area.

3.3 Fact-filtering using Rules

We have applied rules to select extreme values ("superlatives"). Pakistan exports to different countries, but its uniqueness comes from applying the rule max (Export). We first extract and capture the percentage exports from the Factbook dataset (URL7) and then identify the highest export country: Pakistan's largest export partner is the US, while its largest import partner is China.

3.4 Geography Stories

The following are some interesting geographic facts typical of a country:

- Landlocked
- Longest boundary with a particular neighbour
- Largest river
- Rich in a particular resource (diamonds in South Africa, oil in some countries...)

An interesting element is whether country is landlocked or not, for example, Switzerland, or Austria. This fact can be verified by the map provided in stories. Afghanistan is a landlocked country shown in Figure 3.

Although most countries share boundaries with others, it might be interesting to note which country has longest or shortest boundary with the selected country. Afghanistan has as the longest boundary with Pakistan, at approximately 2,670 km. Also, Afghanistan has as the shortest boundary with China (91 km), as shown in Figure 3.

3.5 Unique Facts Identification from Wikipedia Articles using NLP

In this section, we will mention some ideas on how to find relevant and interesting facts using text processing as the introductory part of our story. We have extracted two types of sentences:

1. Sentences with superlatives/adjectives
2. Sentences with years (for capturing historic events).

3.5.1 Identifying Interesting sentences

Interesting sentences about some countries taken from Wikipedia articles are listed below:

Honduras: Honduras has the highest murder rate in the world.

Cuba: In 2015, it became the first country to eradicate mother-to-child transmission of HIV and syphilis, a milestone hailed by the World Health Organization as "one of the greatest public health achievements possible".

Grenada is also known as the "Island of Spice" because of the production of nutmeg and mace crops, of which it is one of the world's largest exporters.

Ghana is one of the world's largest gold and diamond producers, and is projected to be the largest producer of cocoa in the world as of 2015.

Papua New Guinea is one of the world's least explored countries, culturally and geographically, and many undiscovered species of plants and animals are thought to exist in its interior.

3.5.2 Historic year identification

We have identified historic events about countries (e.g. Belarus) by capturing sentences with years such as:

1) Many of the borders of Belarus took their modern shape in 1939 when some lands of the Second Polish Republic were reintegrated after the Soviet invasion of Poland, and were finalized after World War II.
2) In 1945, Belarus became a founding member of the United Nations.

3.6 Food Items

We also list food items in our stories. The site "Food in Every Country" presents the food of some countries with recipes (URL8). We use a list of food items as a starting point and fill our stories with dishes mentioned in the list (URL9). Nasi lemak is the national dish of Malaysia, while tumpeng is considered as the national dish of Indonesia. Kabuli palaw is the favourite dish of

smislu te se smatra da u njoj živi velik broj za sad neotkrivenih vrsta biljaka i životinja.

3.5.2. Utvrđivanje povijesnih godina

Izdvajanjem rečenica s godinama utvrdili smo povijesne događaje u zemljama kao što je Bjelorusija:

1) Većina bjeloruskih granica poprimila je svoj suvremeni oblik 1939., kad je dio zemalja Druge Poljske Republike ponovno pripojen nakon sovjetske invazije Poljske, te su finalizirane nakon Drugog svjetskog rata.
2) Bjelorusija je postala član osnivač Ujedinjenih naroda 1954.

3.6. Hrana

U svojim pričama također navodimo hranu. Stranica *Food in Every Country* (hrana u svakoj zemlji) predstavlja hranu i recepte nekih zemalja (URL8). Počeli smo s popisom hrane i ugradili u naše priče jela s popisa (URL9).

Nasi lemak je nacionalno jelo u Maleziji, dok je Tumpeng nacionalno jelo u Indoneziji. Kabuli Palaw je omiljeno jelo Afganistanaca, a Austrija je poznata po kuhanoj govedini (Tafelspitz) i bečkom odresku (Wiener Schnitzel).

3.7. Povijesni izumi

Povezanost tisuće izuma iz ljudske povijesti sa zemljama nije trivijalna. Pokušali smo povezati popise izuma specifičnih za određenu zemlju. Kvarcni sat izumljen je u Kanadi. Daljinski upravljač pripada Austrougarskoj. Mikroskop i teleskop pripisuju se Nizozemskoj (URL10). Ručni sat dolazi iz Švicarske.

3.8. Povijest lansiranja i proizvodnje vozila

Ljudi su se oduvijek bavili istraživanjem svemira. Ponekad ih zanima povijest vozila. Stoga smo naveli povijest vozila u različitim zemljama. Na primjer, Sovjetski Savez je prva zemlja koja je lansirala Sputnik 1 1957. godine. Prema Wikipediji: "Do prosinca 2013. šezdeset i jedna zemlja lansirala je umjetne satelite". Zanimljivi primjeri vozila su podmornice, kao što je npr. USS Nautilus (SSN-571), prva podmornica na nuklearni pogon na svijetu. German Flocken Elektrowagen iz 1888. smatra se prvim električnim automobilom na svijetu. Kina proizvodi najviše automobila na svijetu i 2015. godine proizvela ih je oko 23 722 890, vidi OICA (URL11).

3.9. Morske luke

Iako su zračni promet i drugi oblici prometa vrlo česti, važan je i promet brodovima. Na primjer, samo se u Europi nalazi 3024 morske luke (URL12). Najveći broj morskih luka u Europi nalazi se u Ujedinjenom Kraljevstvu (731). Treba spomenuti da luka ima i u zemljama koje nemaju izlaz na more jer brodovi mogu putovati velikim rijekama kao što je Dunav u Austriji.

3.10. Radna snaga – prema profesiji

Parametar radne snage pokazuje raspodjelu ljudi koji rade u različitim profesijama. Slijede rečenice izabrane za priče upotrebom pravila max:

- 93,6% ljudi u Brundiju rade kao poljoprivrednici
- 90% ljudi u Burkini Faso rade kao poljoprivrednici
- 90% ljudi u Malaviju rade kao poljoprivrednici.

3.11. Dobitnici Nobelove nagrade

Nobelove nagrade dodjeljuju se ljudima i organizacijama od 1901. (URL13). Izdvojili smo dobitnike Nobelove nagrade (URL13) zajedno s metapodacima (zemlja rođenja, zemlja pripadanja, godina, područje, itd.). Slika 4. prikazuje dob dobitnika Nobelove nagrade rođenih u Austriji. Na primjer, Karl von Frisch dobio je Nobelovu nagradu kad je imao 87 godina.

Gornja se analiza odnosi samo na jednu zemlju. Uključili smo i statističke podatke za druge zemlje. Slijedi nekoliko neobičnih činjenica koje smo pronašli:

- SAD ima najveći broj žena koje su dobile Nobelovu nagradu (11) (brojeći one rođene u SAD-u).
- Šesnaest dobitnika Nobelove nagrade rođeno je u Švicarskoj, ali nijedan u posljednjih 10 godina.
- Deset dobitnika Nobelove nagrade rođeno je u Australiji, ali nijedan od njih nije ju dobio za književnost, mir ili ekonomiju.
- Marie Curie rođena je u Poljskoj, a dobila je Nobelovu nagradu za fiziku i kemiju 1903. i 1911.
- John Bardeen rođen je u SAD-u, a dobio je Nobelovu nagradu za fiziku 1956. i 1972.
- Dvanaest dobitnika Nobelove nagrade rođeno je u Danskoj, a između dviju nagrada bio je razmak od 31 godinu (1944–1975).
- Jedanaest od dvanaest dobitnika Nobelove nagrade rođenih u Kini još je živo.

4. Dodavanje slika i videoisječaka kako bi priče bile zanimljivije

Činjenice smo u pričama proširili smislenim slikama sa stranice PixaBay (URL14) i drugih zbirki slika. Slike

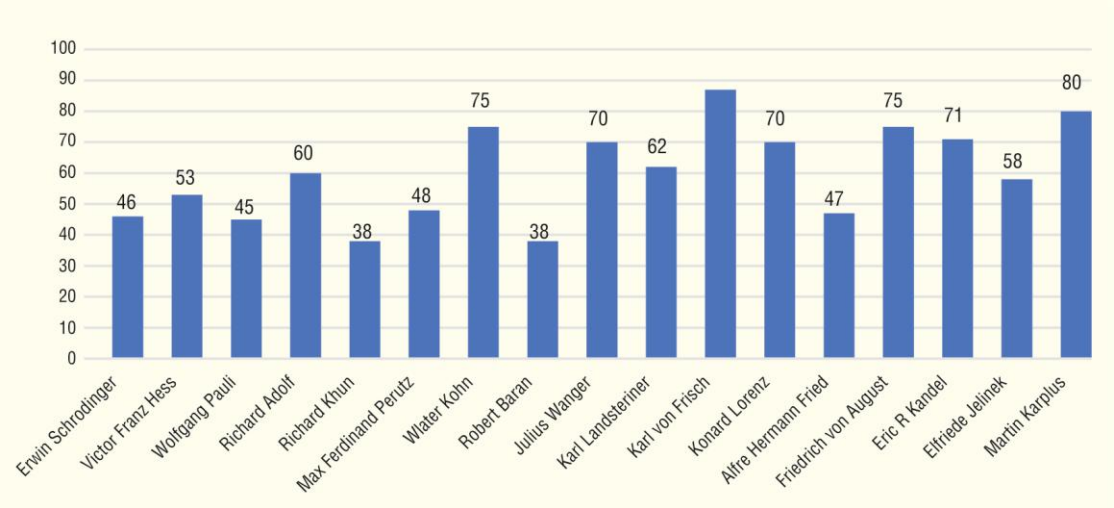

Fig. 4 Nobel Prize winners from Austria
Slika 4. Dobitnici Nobelove nagrade rođeni u Austriji

the Afghan nation. Austria is famous for Tafelspitz and Wiener Schnitzel.

3.7 Historical Inventions

The association of thousands of inventions in human history with countries is far from trivial. We have attempted to integrate country-specific lists of inventions in our story generation section. For example, the quartz watch was invented in Canada. Remote control is an invention that belongs to Austria-Hungary. The microscope and telescope are attributed to The Netherlands (URL10).

The wrist watch is considered an invention from Switzerland.

3.8 Launch History and Production of Vehicles

Exploring the universe is always a human concern. Sometimes, people want to see the launch history of vehicles, so we list them for several countries. For example, the Soviet Union was the first country to launch "Sputnik 1" in 1957. According to Wikipedia, "As of December 2013, sixty-one countries have operated artificial satellites". An interesting example is the USS Nautilus (SSN-571), which was the world's first operational nuclear-powered submarine. The 1888 German Flocken Elektrowagen is regarded as the first electric car in the world. China ranks first in car-producing countries, with about 23,722,890 produced in 2015, see OICA (URL11).

3.9 Seaports

Airports and other transportation means are common, but an important means of communication are seaports. There are 3,024 sea ports (a seaport is a facility which can accommodate ships which go out to sea) in Europe (URL12). Considering Europe, the largest number of seaports is in the United Kingdom (731). Note that seaports can also exist in landlocked countries, because a large river allows seagoing vessels to navigate. The River Danube in Austria is one such example.

3.10 Labour Force – by Occupation

The labour force parameter shows the distribution of people working in different professions. The following are sentences picked for stories using the max rule:

- 93.6% of people are engaged in the occupation "Agriculture" in Burundi
- 90% of people are engaged in the occupation "Agriculture" in Burkina Faso
- 90% of people are engaged in the occupation "Agriculture" in Malawi

3.11 Nobel Prize Winners

Since 1901, Nobel Prizes have been awarded to different people and organizations (URL13). We have extracted Nobel Prize-winner lists from (URL13) along with the meta data (country of birth, country of affili-

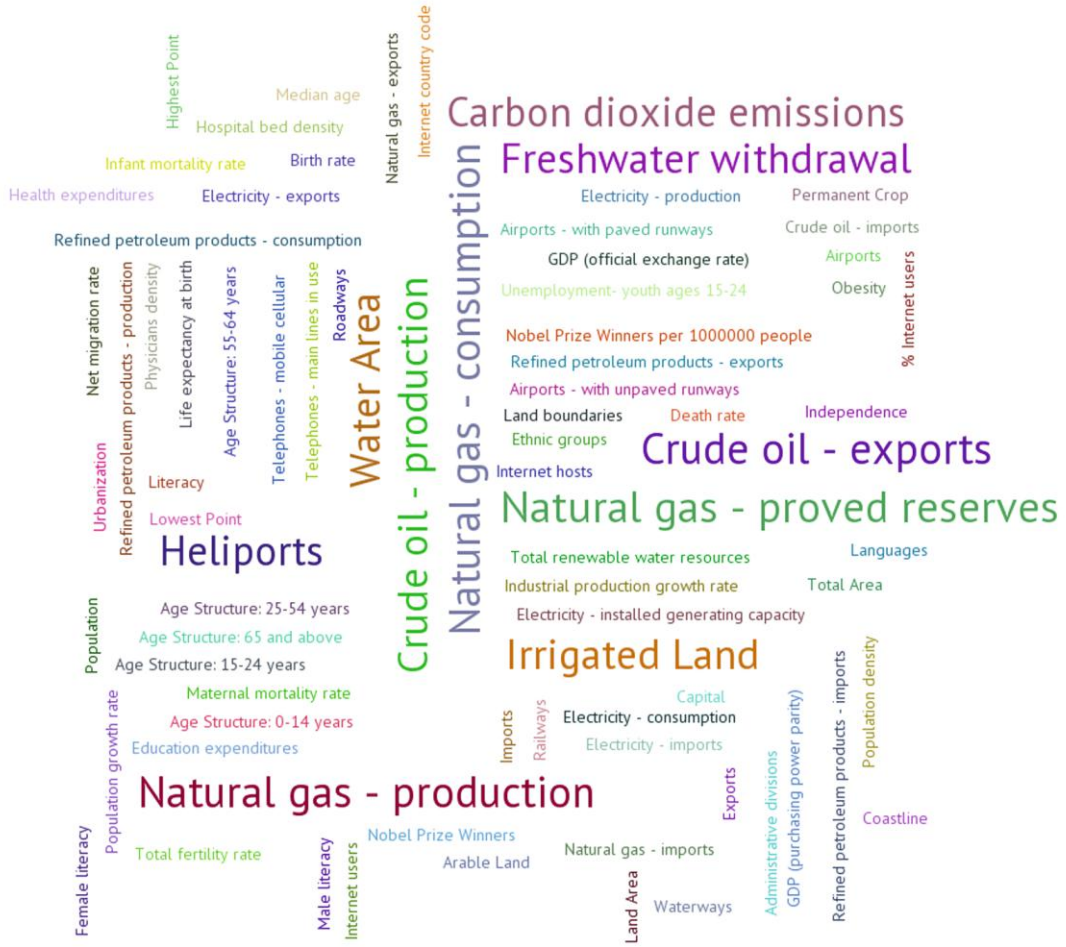

Fig. 5 Tag cloud representing various properties for Iran
Slika 5. Oblak riječi koji prikazuje različita svojstva Irana

smo preuzeli i s Austria-Foruma, posebno global-geography.org, Wikipedia, Flickr, Factbook, itd. Upotrijebili smo i poveznice na videoisječke na stranici YouTube, što se vidi u priči o Mađarskoj (URL15). Također smo upotrijebili slike iz skupa podataka YFCC100M (The New Data in Multimedia Research) (Bard i dr. 2016). YFCC100M najveća je zbirka multimedije preuzeta sa stranica Flickr, Instagram i Yahoo.

5. Oblaci riječi

Oblak riječi (tag cloud, word cloud) vizualni je prikaz tekstualnih podataka. Obično se sastoji od pojedinačnih riječi koje se razlikuju po veličini ili boji. Izradili smo oblake podataka koji prikazuju zemlje svojstvima po kojima su te zemlje visoko rangirane. Oblak podataka za Iran prikazan je na slici 5. Iran se nalazi na visokom mjestu po svojstvima kao što su zalihe prirodnog plina, površina vode, proizvodnja sirove nafte, itd. No u Iranu je također velika emisija ugljikovog dioksida.

6. Zaključak i budući radovi

U uvodnom dijelu ovoga rada naveli smo niz primjera koji su opisali kako smo došli do „zanimljivih činjenica“ upotrebom pravila (heuristika).

Te su nam heuristike pomogle stvoriti priče o zemljama. One se mogu spojiti u prihvatljiv oblik bez mnogo ljudske intervencije kad se upotrijebi ograničen broj baza podataka o određenoj temi (ovdje: istaknutosti zemalja svijeta). Međutim, smatramo da se slični postupci mogu primijeniti u bilo kojem području, što može u potpunosti promijeniti način na koji pretražujemo internet. Kako bi to bilo moguće, potrebno je poredati baze podataka po kvaliteti, kao što to možemo učiniti s restoranima. Jasno, baze podataka najveće kvalitete bit će one koje se najviše upotrebljavaju. Potom je potrebno ukloniti informacije koje se pojavljuju više puta, što zahtijeva prilično složene tehnike jezične obrade. Rezultati će biti stavljeni u izvještaj u kojem će pisati odakle su podatci došli, što će omogućiti korisnicima da po želji istraže određeni izvor. Držimo da taj izvještaj treba

ation, year, field etc.). Figure 4 shows the ages of different Nobel Prize winners from Austria (born in Austria). Karl von Frisch got the Nobel Prize at the ripe old age of 87.

The analysis above is based on only one country. We went further and involved other country statistics for comparison. Some of the many unusual facts we found are:

- The United States has the largest number of women Nobel Prize winners (11) (including women born in the United States)
- There have been 16 Swiss-born Nobel Prize winners, but none from that country in the last ten years.
- There are 10 Australian-born Nobel Prize winners, but none in literature, peace, or economics.
- Polish-born Maria curie got the Nobel Prize twice in two different categories (physics and chemistry) in 1903 and 1911.
- John Bardeen, born in the United States, got the Nobel Prize for physics twice, in 1956 and 1972.
- There are 12 Danish-born Nobel Prize winners, but there was a gap of 31 years (1944-1975) between two awards.
- Of the 12 Chinese-born Nobel Prize winners, 11 are still alive today.

4 Adding Pictures and Videos to Make Stories More Interesting

Besides factual information, we are extending our stories using meaningful pictures from PixaBay (URL14) and other picture collections. Other sources include all the pictures in Austria- Forum, particularly global-geography.org, Wikipedia, Flickr, Factbook etc. Further, we are using links to videos from YouTube in the appropriate sections; see the story of Hungary (URL15). We are also using pictures from the dataset YFCC100M (The New Data in Multimedia Research), see (Bard et al. 2016). The YFCC100M is the largest multimedia collection ever released, taken from Flickr, Instagram and Yahoo.

5 Tag Clouds

A tag cloud is a visual representation of textual data. Tags are usually single words. The importance of each tag is shown by font size or colour. We have created tag clouds representing countries using the properties in which they rank highly. The tag cloud for Iran is shown in Figure 5. Iran ranks highly in many properties, such as proven natural gas reserves, water area, crude oil production, etc. But also, CO2 emission is high in Iran.

6 Conclusion and Future Work

To give an idea of how we try to discover "interesting facts" by means of a set of rules (heuristics), we have presented a set of examples in Section 1: Introduction.

These heuristics helped us generate stories about countries. The stories obtained can be combined in an acceptable format without substantial human intervention, when a limited set of databases on a particular topic (here: highlights of countries of the world) is used. However, it is our vision to apply similar techniques to any domain, changing how we search the Web thoroughly. To make this feasible, it will be necessary to rank databases by quality, as we do today with restaurants. Clearly, the highest quality properties will be queried most. Then, from the large set of pages obtained from queries, duplicate information will be removed, requiring quite sophisticated language processing techniques. The results will be compiled in a report, with each section showing where it came from, i.e. allowing the user to explore a particular site if desired. We believe that the report should be created dynamically (avoiding questions of copyright infringements). For the foreseeable future, the reports will be somewhat unstructured, yet reading such a report instead of looking at numerous Web sites will save quite a lot of time. Further, if the domain and set of servers is small enough, then with limited human intervention, it will be possible to generate coherent stories, as we have tried to show in the project described in this paper.

To be quite immodest, we believe this will revolutionize searching and make systematic access to information on the Web much more effective, yet some major efforts will still be necessary to reach this aim. In the foreseeable future, small further steps along the ones initiated by us will make the heuristic combination of material from various sources an increasingly serious alternative.

izraditi na dinamičan način (što izbjegava probleme autorskih prava). Ti će izvještaji do daljnjega biti donekle nestrukturirani, no čitanje takvih izvještaja umjesto mnogobrojnih web-stranica uštedjet će dosta vremena. Nadalje, ako su područje i skup poslužitelja dovoljno mali, tada će s ograničenim ljudskim intervencijama biti moguće stvoriti koherentne priče, kao što smo pokušali pokazati u projektu opisanom u ovom radu.

Neskromni smo, ali smatramo da će naša metoda revolucionirati pretraživanje interneta i učiniti ga mnogo učinkovitijim. U predvidivoj budućnosti bit će potrebni još neki mali koraci kako bi heurističko povezivanje informacija iz različitih izvora postalo ozbiljnijom alternativom.

References / Literatura

Bard T, David S, Friedland G, Elizalde B, Karl N, Poland D, Borth D, Jia Li L, 2016, vfcc100M: The New Data in Multimedia Research, Commun. ACM, 59 (2), pp 64–73

Maurer H, Mehmood R, 2015, Merging image databases as an example for information integration, CEJOR, 23(2), pp 441–458

Medelyan O, Milne D, Legg C, Witten I, 2004, Mining meaning from Wikipedia, International Journal of Human-Computer Studies, 67(9), pp 1–76

Mehmood R, Kulathuramaiyer N, Maurer H, 2016, A New Look at Geography of the World, IPSI, 12(1), pp 21–29

Nakayama K, Hara T, Nishio S, 2008, Wikipedia link structure and text mining for semantic relation extraction towards a huge scale global web ontology, CEUR Workshop Proc., vol. 334, pp 59–73

Yeung C A, Jatowt A, 2011, Studying How the Past is Remembered: Towards Computational History through Large Scale Text Mining, Proc. 20th ACM Int. Conf. Inf. Knowl. Manag., pp 1231–1240

URL 1: Austria Forum, https://austria-forum.org/af/Geography (October 1, 2016)

URL 2: Austria Forum, https://austria-forum.org/af/Geography/Main_Ideas/Current_List_of_Stories (September 15, 2016)

URL 3: Maps of World, http://www.mapsofworld.com/world-top-ten/most-patent-registering-countries.html (October 1, 2016)

URL 4: Touropia, http://www.touropia.com/famous-trees-in-the-world/ (September 15, 2016)

URL 5: The Flower Expert, http://www.theflowerexpert.com/content/aboutflowers/national-flowers (September 15, 2016)

URL 6: List 25, http://list25.com/25-longest-bridges-in-the-world/5/ (September 15, 2016)

URL 7: Jmatchparser, http://jmatchparser.sourceforge.net/factbook/ (October 1, 2016)

URL 8: Food by Country, http://www.foodbycountry.com/ (September 15, 2016)

URL 9: Wikipedia, https://en.wikipedia.org/wiki/National_dish (September 15, 2016)

URL 10: Eupedia, http://www.eupedia.com/europe/list_of_inventions_by_country.shtml (October 1, 2016)

URL 11: OICA, http://www.oica.net/category/production-statistics/ (October 1, 2016)

URL 12: Ports, http://ports.com/ (September 15, 2016)

URL 13: Nobel Prize, http://www.nobelprize.org/ (September 15, 2016)

URL 14: Pixabay, https://pixabay.com/en/edelweiss-leontopodium-microdochium-818414/ (October 1, 2016)

URL 15: Austria Forum, https://austriaforum.org/af/Geography/Europe/Hungary/Special_Information/Surprising_Facts (October 1, 2016)

URL 16: OpenLayers, http://openlayers.org/ (October 1, 2016)

URL 17: Leaflet, http://leafletjs.com/ (October 1, 2016)

URL 18: Carto DB, https://cartodb.com/ (October 1, 2016)

URL 19: CIA World Factbook, https://www.cia.gov/library/publications/resources/the-world-factbook/ (September 15, 2016)

URL 20: National Geographic, http://travel.nationalgeographic.com/travel/countries/pakistan-guide/ (September 15, 2016)

URL 21: Natural Earth Data, http://www.naturalearthdata.com/ (September 15, 2016)