

CLARIN

Common Language Resources and Technology Infrastructure



Projekt CLARIN i CLARIN ERIC Europske istraživačke infrastrukture za humanističke i društvene znanosti

Marko Tadić

Odsjek za lingvistiku
Filozofski fakultet
Sveučilište u Zagrebu
marko.tadic@ffzg.hr

Okrugli stol projekta SERSCIDA
Zagreb
2012-11-14

Što je istraživačka infrastruktura?



2012-11-14

- različita moguća značenja naziva “istraživačka infrastruktura”
 - mreža agencija za financijsku potporu istraživanja
 - mreža ustanova (instituti, sveučilišta/fakulteti, zavodi itd.)
 - mreža stručnih i znanstvenih udruga
 - mreža uređaja koji omogućuju istraživanja
 - npr. Veliki sudarač hadrona, razni ubrzivači, teleskopi...
 - mreža arhiva u kojima su pohranjeni rezultati istraživanja
 - digitalne knjižnice (objavljena djela)
 - inicijativa za stalni pristup (inicijativa *Open Access*, ESF)
 - mreža arhiva u kojima su pohranjeni primarni istraživački podatci i sredstva za pristup njima
 - inicijativa za stalni pristup (inicijativa *Permanent Access*, ESF)
- zadnje značenje bitno za humanističke i društvene znanosti (HiDZ)

Što je istraživačka infrastruktura?



2012-11-14

- iza ovoga zadnjega značenja stoji ozbiljna epistemološka postavka: tekst je iznimno bitan za HiDZ
 - objekt istraživanja tih znanosti: sâm tekst (jezik/govor tog teksta)
 - lingvistika, sve (neo)filologije, fonetika itd.
 - objekt istraživanja tih znanosti: posredovan tekstem
 - teorija književnosti, povijest, sociologija, psihologija, kognitivna znanost, umjetna inteligencija, itd.
- količina istraživačkih podataka u HiDZ u obliku e-teksta raste nevjerojatnom brzinom
 - masovnom digitalizacijom tradicionalnih tekstnih medija
 - stvaranjem novoga e-teksta (bez papira)
- (računalna) lingvistika (tj. formalni pristupi jeziku)
 - za HiDZ = matematika kad je riječ o (e-)tekstu
 - tj. ima ulogu koju matematika ima među prirodnim znanostima
- jezične tehnologije = tehnička primjena tih znanja

Problemi u eHiDZ



- masovna digitalizacija istraživačkih podataka
- većina se digitalnih arhiva temelji na tekstovnim dokumentima
 - ponajprije jezično, ali također i slikovno ili zvukovno utemeljeni
- mnogi arhivi poznati samo lokalnim korisnicima
- arhivi su mahom nepovezani
 - uz nekoliko iznimaka (DOBES/MPI, DELAMAN projekt itd.)
- svaki arhiv rabi svoj zaseban standard za pohranu i pristup
 - najčešće jednostavno pretraživanje datoteka (Ctrl-F, F3...)
- istraživači u HiDZ najčešće nisu svjesni mogućih prednosti korištenja jezičnih resursa i alata za pristup e-tekstu
- te resurse i alate teško koriste ne-specijalisti
- istraživači u HiDZ najčešće se ne razumiju u JT

Misija projekta CLARIN



2012-11-14

- što?
 - stvoriti istraživačku infrastrukturu koja omogućuje uporabu jezičnih resursa i tehnologija istraživačima ih svih znanosti, a osobito HiDZ
- kako?
 - ujediniti postojeće digitalne arhive u **europsku federaciju digitalnih arhiva**
 - ciljana skupina korisnika: **znanstvenici iz HiDZ**
 - pristup: **jedinstven i jednoobrazan pristup** svim arhivima
 - sredstva: **jezičnotehnološki alati** kao web-servisi za pretragu, rukovanje, obogaćivanje i korištenje podataka iz tih arhiva
 - pokrivenost: **svi su jezici** jednako bitni
 - razmještaj: uključene **sve države EU i države kandidati**

Zašto europska infrastruktura?



2012-11-14

- previše udrobljenosti po zemljama
 - manjak koordinacije
 - manjak vidljivosti
 - manjak suradnje
 - manjak održivosti
- ekspertiza postoji, ali ne u svim zemljama
- jezično neovisni alati mogu se razmjenjivati
- jezično ovisni alati često se mogu prilagoditi za druge jezike
- većina zemalja ne može sama podnijeti cijenu izgradnje takve infrastrukture

Zašto sada?



- eksponencijalni rast digitalnih podataka
- zrelost jezičnih tehnologija
 - omogućuju brzu obradu podataka (GRID računarstvo)
 - omogućuju obradu velikih količina podataka (GRID)
 - omogućuju višejezičnu obradu i pristup (e-znanost)
 - omogućuju nova istraživačka pitanja (e-znanost)
- rastuće zanimanje na razini EU-a za istraživačke infrastrukture u Europskome istraživačkome prostoru (ERA)
- dopunjen ESFRI Roadmap (2006, 2008, 2010) uključuje 48 prijedloga za istraživačke infrastrukture
 - HiDZ: CLARIN, DARIAH, CESSDA, ESS, SHARE
- *European RI Consortium* (ERIC)
 - zakonski okvir za nadnacionalne EU istraživačke infrastrukture

Što je bio projekt CLARIN?



- CLARIN = kratica za
 - Common Language Resources and Technology Infrastructure (for Humanities and Social Sciences)
- projekt 7. okvirnoga programa, Istraživačke infrastrukture
 - 2008-01-01 – 2011-06-30
- izgradnja istraživačke infrastrukture u 3 faze
 - pripremna faza
 - 2008.-2011.: planiranje, izgradnja prototipa
 - proračun: 4.1 M€ od EK, oko 5 M€ od sudjelujućih zemalja
 - gradbena faza
 - 2011.-2013.: izgradnja i popuna resursima i alatima
 - uporabna faza
 - 2014.-: CLARIN infrastruktura u punome pogonu
- ukupan proračun (2008.-2018.): očekivano oko 113 M€

Što je bio projekt CLARIN?



2012-11-14

- konzorcij CLARIN-a
 - 33 partnera iz 22 EU zemlje-članice i zemalja-pristupnica
- zajednica CLARIN-a
 - 151 član iz 33 zemlje
 - zanimanje i iz Amerika, Azije, Australije
- vodeći partneri
 - Sveučilište u Utrechtu (*Steven Krauwer, koordinator*)
 - Max Planck Institut, Nijmegen (*Peter Wittenburg*)
 - Mađarska akademija znanosti (*Tamás Váradi*)
 - Sveučilište u Oxfordu (*Martin Wynne*)
 - ...
- CLARIN je utemeljen na četiri prethodne inicijative
 - LangWeb, EARL, TELRI I. i II., DAM-LR

Što je CLARIN ERIC?



2012-11-14

- osnovan 2012-04-18
- 9 (zemalja)-osnivačica
 - Austrija, Bugarska, Češka, Danska, Estonija, Nizozemska, Nizozemska jezična unija, Njemačka, Poljska
 - Hrvatska nije među njima premda je potpisala Memorandum o razumijevanju
 - MVEP nije dostavilo potvrdu da RH priznaje ERIC kao nadnacionalnu pravnu osobu
 - 5-godišnja članarina za RH: 61.408 EUR
- svaka članica preuzela je obvezu financiranja nacionalnoga CLARIN konzorcija
 - izgrađuje infrastrukturu na nacionalnoj razini
 - povezuje tu infrastrukturu s CLARIN ERIC-om na EU razini
- www.clarin.eu



- zahvaljujući razvijenoj istraživačkoj infrastrukturi istraživači iz HiDZ moći će bez poteškoća postavljati ovakva pitanja:
 - “Popiši sve uporabe riječi entuzijizam u 19-stoljetnim engleskim romanima koje su napisale žene!”
 - “Prikaži sve zemljovide između 15. i 18. stoljeća koji pokazuju Dravu i Osijek!”
 - “Pokaži mi sve moguće prijevode riječi saudade iz djela Camõesa i koje su se od njih zaista pojavile u prijevodima njegovih djela na litavski!”
 - “Tko se sve pojavljuje u istim dokumentima s Nikolom Šubićem Zrinjskim u 18. st.?”
 - “Sažmi *Le Monde* od 12. studenoga 2009. na bugarskome!”
 - primjer: puni sustav za prepoznavanje imena u www-izdanju dnevnih novina na hrvatskome →

Traži

17. veljače 2005.

Home

Aktualnosti

Vijesti

Crna kronika

Sport

Scena

Kultura

Gospodarstvo

→ **Poslovni svijet**

Zanimljivosti

Regije

Free Time

Kompas

Događanja

Kino

TV Vodič

Vrijeme

Lifestyle

Nedjeljni Večernji

NOVO

Vijesti bez slika

Moja karijera

e-Shop

→ Gospodarstvo → **Poslovni svijet**

25.01.2005 18:45

**Hrvatski izvoz još na niskim razinama
90 posto tvrtki uopće ne izvozi!**Autor **Piše Josip Bohutinski**

Hrvatski izvoz napokon je prošle godine počeo rasti brže od uvoza te je, prema podacima za prvih 11 mjeseci 2004. godine, izvoz u kunama rastao 15,7 posto a uvoz 5,7 posto. Iz Hrvatske je izvezeno robe u vrijednosti nešto manjoj od 44 milijardi kuna ili 7,25 milijardi američkih dolara, dok je vrijednost uvoza bila 91,19 milijardi kuna ili više od 15 milijardi dolara.

No podaci o izvozu po glavi stanovnika upozoravaju da je hrvatski izvoz još na niskim razinama u usporedbi s drugim i sličnim zemljama. Prema podacima udruge Hrvatski izvoznici, u 2003. godini vrijednost hrvatskog izvoza po glavi stanovnika bila je samo 1106 dolara.

Koliko je je to mala vrijednost, govori podatak o slovenskom izvozu po glavi stanovnika od čak 4774 dolara. Irska na svakog svoga stanovnika izveze 22.119 dolara roba i usluga. Amerikanci, pak, po glavi stanovnika izvezu robe u vrijednosti 2360

→ **Ostale vijesti**

- ▶ Novi igrači zajedno protiv T-HT-a?
- ▶ Hrvatska ipak nije prezadužena
- ▶ Povjerenstva odvaguju veletrgovine
- ▶ Državne banke kreću u investicijsko bankarstvo
- ▶ Dubrovnik ne bi smio dominirati u promidžbi
- ▶ Fokus na Sirelu i Sombled
- ▶ Dioničari ne žele novac nego banku

Prijava

Korisničko ime

Lozinka

Prijava

- Registrirajte se!
- Zaboravili ste lozinku?

Kolumna

→ Nakon revolucije,
slijedi evolucija

→ Devizno tržište

→ Tržište novca

→ Tržište obveznica

→ Tržište kapitala

→ Karijere

moja
karijera

Večernji list, 2005-02-17, Gospodarstvo Hrvatski izvoz još na niskim razinama **90 posto tvrtki uopće ne izvozi!**

Autor Piše *Josip Bohutinski*

Hrvatski izvoz napokon je **prošle godine** počeo rasti brže od uvoza te je, prema podacima za **prvih 11 mjeseci 2004. godine**, izvoz u kunama rastao **15,7 posto** a uvoz **5,7 posto**. Iz **Hrvatske** je izvezeno robe u vrijednosti nešto manjoj od **44 milijardi kuna** ili **7,25 milijardi američkih dolara**, dok je vrijednost uvoza bila **91,19 milijardi kuna** ili više od **15 milijardi dolara**.

No podaci o izvozu po glavi stanovnika upozoravaju da je hrvatski izvoz još na niskim razinama u usporedbi s drugim i sličnim zemljama. Prema podacima udruge Hrvatski izvoznici, u **2003. godini** vrijednost hrvatskog izvoza po glavi stanovnika bila je samo **1106 dolara**.

Koliko je je to mala vrijednost, govori podatak o slovenskom izvozu po glavi stanovnika od čak **4774 dolara**. **Irska** na svakog svoga stanovnika izveze **22.119 dolara** roba i usluga. Amerikanci, pak, po glavi stanovnika izvezu robe u vrijednosti **2360 dolara**.

No vrijednost izvoza velikih zemalja po glavi stanovnika u pravilu je manja od izvoza malih zemalja zbog velikog domaćeg tržišta koje može apsorbirati veliki dio domaće proizvodnje. To potvrđuju i podaci o izvozu po stanovniku i "malih zemalja" poput **Belgije, Nizozemske i Finske**.

Uz malu vrijednost izvoza po glavi stanovnika, za **Hrvatsku** je nepovoljan i podatak o broju domaćih tvrtki čija godišnja vrijednost izvoza premašuje **milijun kuna**.

Njih je samo **pet posto** od ukupno aktivnih poduzeća. Naime, prema podacima Hrvatskih izvoznika, od 70-ak tisuća aktivnih kompanija u **Hrvatskoj**, svoje proizvode i usluge na strana tržišta izvozi samo njih 6700. Pritom je izvoznika čija vrijednost izvoza premašuje **milijun kuna** samo 3144. Ta grupa izvoznika, prema podacima udruge Hrvatski izvoznici, ostvaruje čak **96 posto** ukupnog hrvatskog izvoza.

Koliko je bitna uloga izvoznika u cjelokupnom hrvatskom gospodarstvu, potvrđuje podatak da 2688 izvoznika izdvaja **83 posto** ukupne dobiti u **Hrvatskoj**, odnosno 16,6 od **19,9 milijardi dolara**.

Upozoravajući na podatke o hrvatskom izvozu po glavi stanovnika, predsjednik Hrvatskih izvoznika **Darinko Bago**, prilikom prošlotjednog potpisivanja Sporazuma o suradnji s **Hrvatskom** bankom za obnovu i razvitak, najavio je sklapanje sličnih sporazuma s drugim udruženjima i institucijama koje mogu pridonijeti afirmaciji hrvatskog izvoza, bez kojeg, naglasio je **Bago**, **Hrvatska** nema budućnosti.

A velike zasluge za prošlogodišnji brži rast hrvatskog izvoza sigurno ima upravo **HBOR** i njegovi programi poticanja izvoza. Preko programa Kreditiranje priprema roba za izvoz i izvoza roba **lani** je odobreno 170 kredita u vrijednosti **1,25 milijardi kuna**, što je čak **448 posto** veći iznos nego **2003. godine** kada su odobrena 52 kredita, ukupno vrijedna nešto više od **279 milijuna kuna**.

I Program osiguranja izvoza zabilježio je **lani** veliki rast. U **2004. godini** osiguran je promet od **580 milijuna kuna**, što je povećanje **180 posto** prema **prethodnoj godini**, a odobreno je 357 zahtjeva, što je povećanje od **306 posto**. **Lani** je **HBOR** osigurao izvoz 67 izvoznika, za razliku od 35 u **2003. godini**. Od početka poslovanja **HBOR** je dosad isplatio 12 odšteta u iznosu **3,2 milijuna kuna**, a od toga je **lani** četvero izvoznika dobilo odštetu od **538.000 kuna**.

Predsjednik Uprave **HBOR-a Anton Kovačev**, potpisujući sporazum s Hrvatskim izvoznicima, rekao je da je **2004.** bila godina izvoza za njegovu banku te da se nada da će ova biti izvozna za cijelu **Hrvatsku**, čemu bi trebao pridonijeti i sporazum o suradnji **HBOR-a i HIZ-a**.

Kovačev je upozorio i da rast hrvatskog izvoza **lani** nije isključivo rezultat brodogradnje.

- Oko **90 posto** kredita koje smo dali za pripremu roba za izvoz i izvoz roba odnosi se na prerađivačku industriju, poput prehrambene, metalske, farmaceutske i drvne industrije. A te industrije su ostvarile porast izvoza **6,5 posto**, što je veći rast od prosječnog ukupnog rasta od **15,7 posto** - rekao je **Kovačev**.

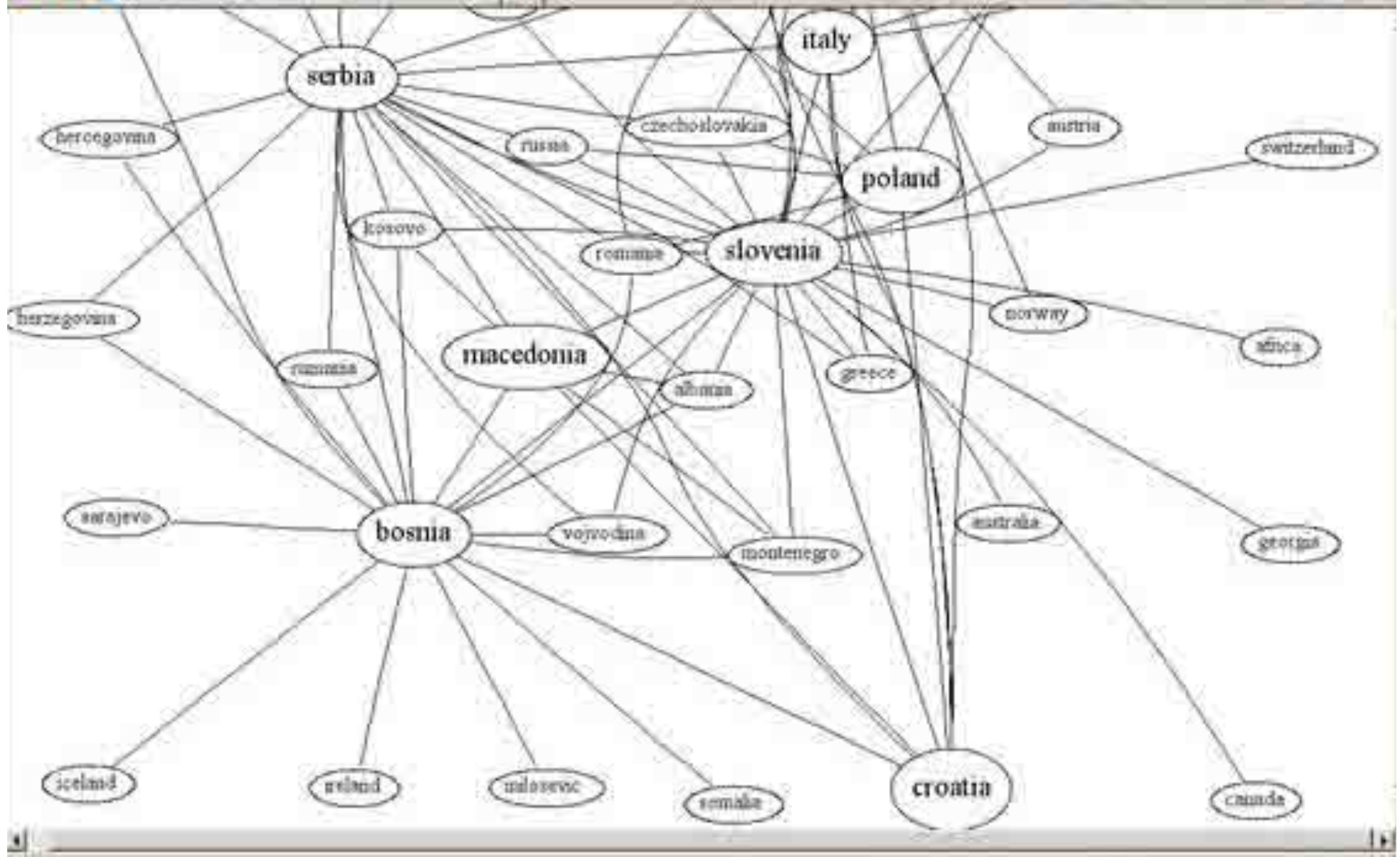
numerical and percentual values

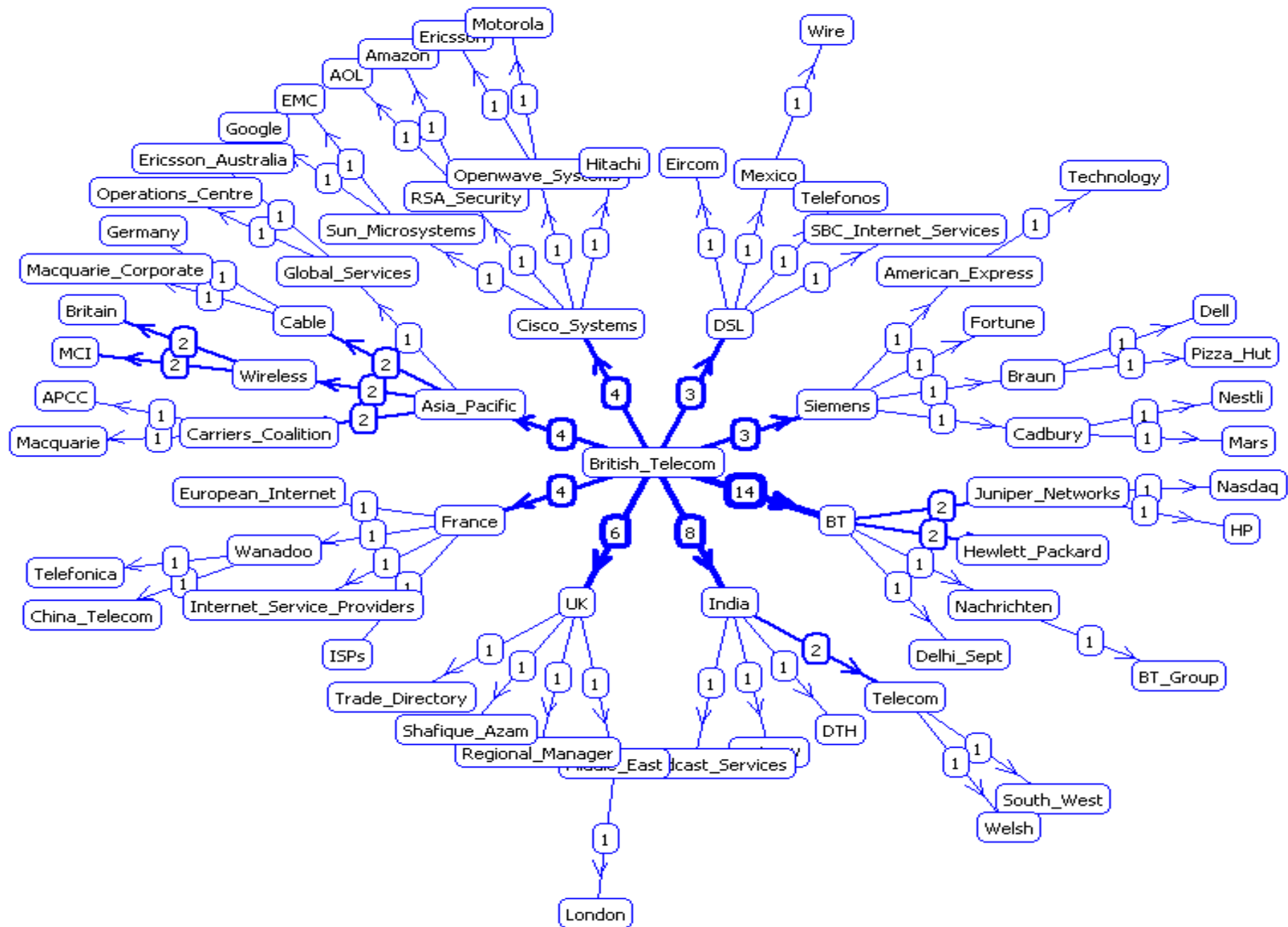
temporal expressions

persons

locations

organizations

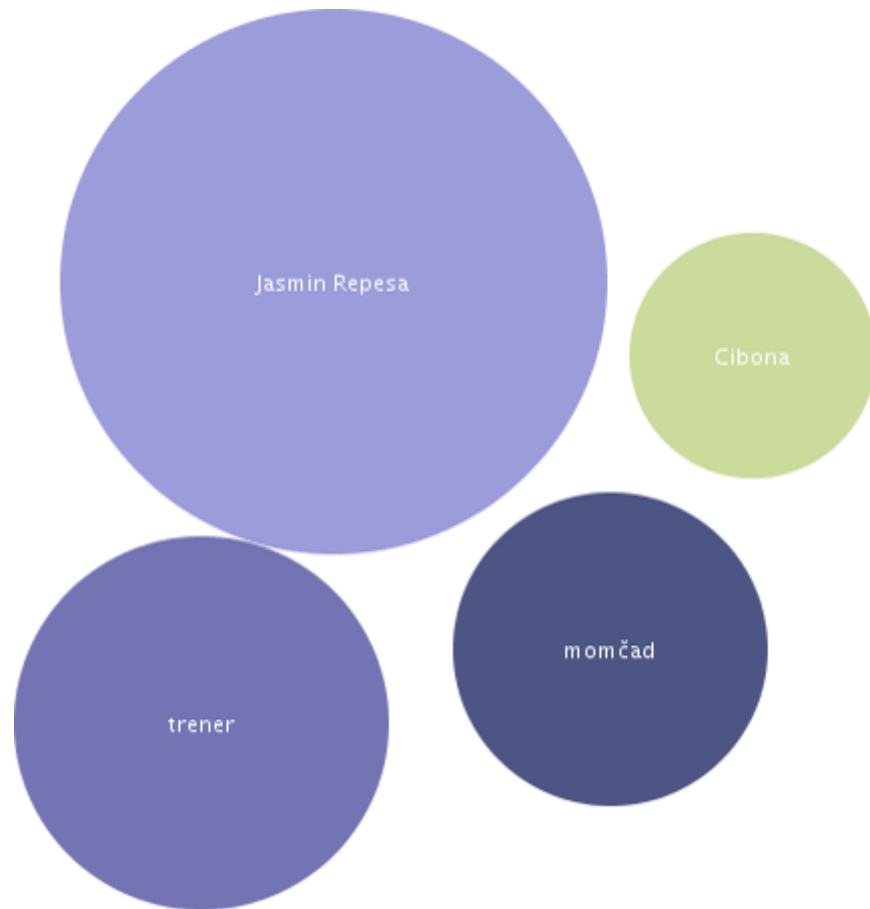




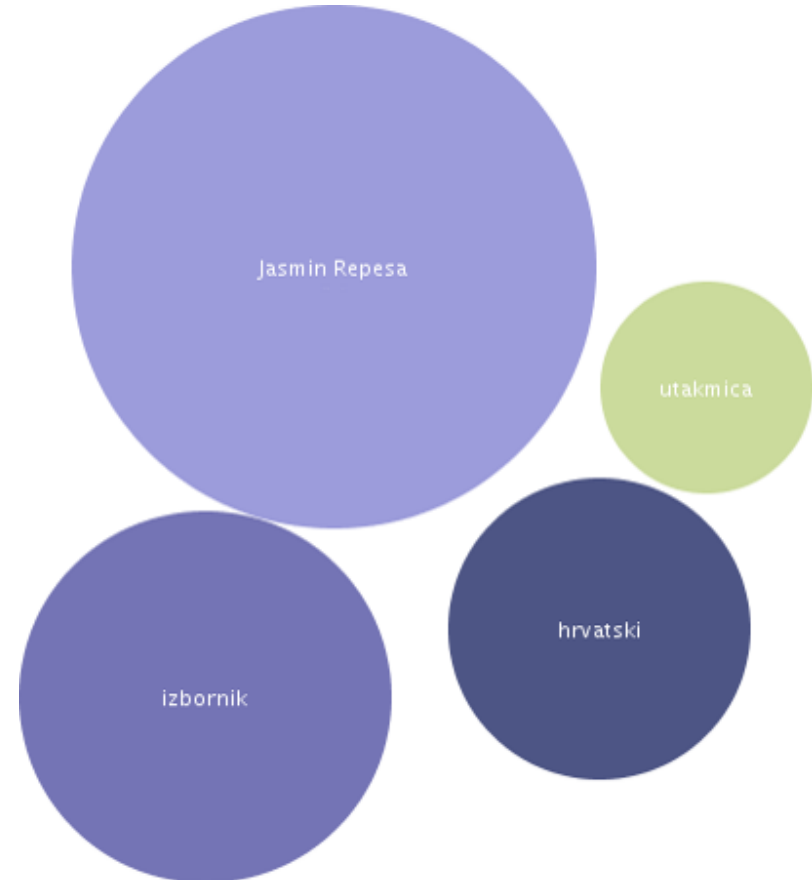
Praćenje imena kroz vrijeme



2001.



2009.



Istraživačke infrastrukture u HiDZ



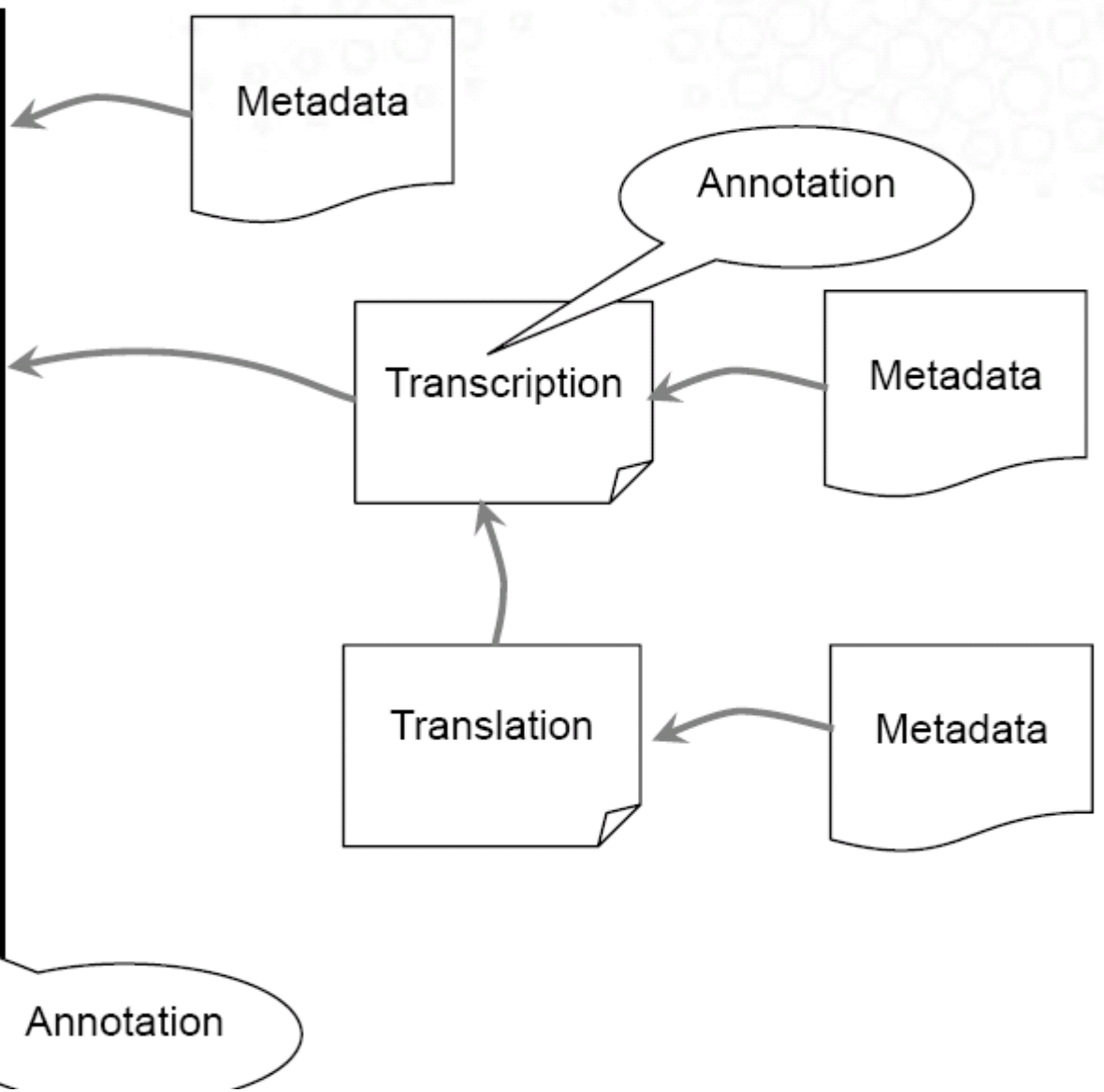
2012-11-14

- kako pospješiti izgradnju istraživačkih infrastrukture u HiDZ?
 - u prirodnim znanostima: ubrzivači, reaktori, brodovi,...
- u HiDZ paradigma e-znanosti = ključna
- omogućuje istraživačima iz HiDZ
 - do sada nezamislive oblike obradbe istraživačke građe

Istraživačke infrastrukture u HiDZ



2012-11-14



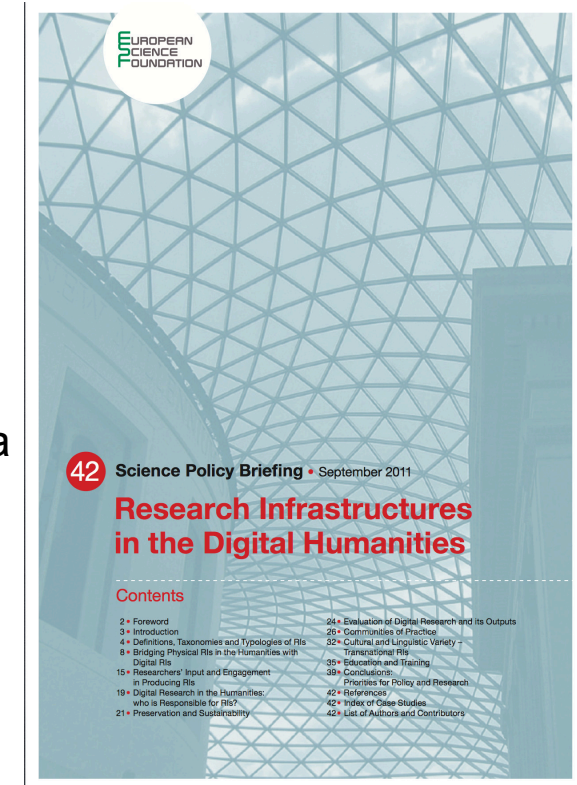


- kako pospješiti izgradnju HiDZ istraživačkih infrastrukture?
 - u prirodnim znanostima: ubrzivači, reaktori, brodovi,...
- u HiDZ paradigma e-znanosti = ključna
- omogućuje istraživačima iz HiDZ
 - do sada nezamislive oblike obradbe istraživačke građe
 - uvid u odnose među digital(izira)nim entitetima koji su do sada bili nemogući
 - izgradnju računalnih modela pojedinih (pod)sustava
 - npr. diskursni odnosi, odnosi među imenima → društvene mreže u zbirkama tekstova
 - hvatanje koraka u HiDZ s paradigmom e-znanosti
- istraživačke infrastrukture u HiDZ više su nego nužne



Goruća pitanja znanstvene politike za istraživačke infrastrukture u humanističkim znanostima (SPB42)

- ESF Science Policy Briefing
 - iskazuju stav ESF-a o gorućim pitanjima znanstvene politike na razini čitave Europe
 - predlažu smjerove znanstvene politike
- SPB su namijenjeni svim ključnim čimbenicima
 - ESF-u i organizacijama-članicama, vladama, EK, EP, međunarodnim agencijama, industriji, akademskim krugovima
- u SPB42 šest osnovnih područja pokriveno
 - snimak stanja i potreba
 - povezivanje fizičkih i digitalnih istraživačkih infrastrukture
 - odabir strateških pravaca
 - podrška za suradnju i umrežavanja
 - vrjednovanje i priznanje rada uloženoga u izgradnju istraživačkih infrastrukture
 - diseminacija





Snimak stanja i potreba

- detaljan inventar **tekućih istraživačkih aktivnosti i buduće potrebe** s obzirom na infrastrukturnu potoporu (npr. standardi, alati, licencije itd.)
 - projekt MERIL (www.meril.eu)
 - jedinstven popis II u Europi za sva znanstvena područja
 - definira kategorije potrebne za opis, definiciju i izgradnju istraživačkih infrastruktura
 - za svako pojedino znanstveno područje (npr. zbirke, arhivi, resursi, pogoni...)
 - transverzalne kategorije (npr. konceptualni modeli, GIS podatci, vremenska crta, vizualizacija...)
- podrška **partnerstvima** između različitih istraživačkih zajednica i ustanova = bolja koordinacija financiranja i aktivnosti na europskoj razini koja uključuje
 - akademsku zajednicu, knjižnice, arhive, muzeje, privatne ustanove
 - javna tijela, komercijalni sektor
- **ESF** se ujedinila s **EUROHORC-om** u novu europsku znanstvenu agenciju **Science Europe**
- članice su države koje imaju razrađen sustav javno financiranih znanstvenih projekata
 - Hrvatske u Science Europe **nema** jer nema ustaljen sustav projekata



Poziv na Dan jezičnih tehnologija

- ❑ Dan jezičnih tehnologija
- ❑ 30. studenoga 2012., 9:00 – 18:00
- ❑ Zagreb, Hotel Sheraton
- ❑ pod visokim pokroviteljstvom Predsjednika Republike Hrvatske
prof. dr. Ive Josipovića
- ❑ sudionici
 - stručnjaci iz
 - Europske komisije
 - mreže izvrsnosti META-NET
 - partnerskih ustanova projekta CESAR
 - hrvatskih ustanova koje se bave jezičnim tehnologijama
 - industrijski partneri
 - političari
- ❑ www.cesar-project.net/hr-roadshow

CLARIN

Common Language Resources and Technology Infrastructure



Projekt CLARIN i CLARIN ERIC Europske istraživačke infrastrukture za humanističke i društvene znanosti

Marko Tadić

Odsjek za lingvistiku
Filozofski fakultet
Sveučilište u Zagrebu
marko.tadic@ffzg.hr

Okrugli stol projekta SERSCIDA
Zagreb
2012-11-14