

# Baza Broadcast News za slovenski jezik (BNSI) in sistem za razpoznavanje tekočega govora

Andrej Žgank, Tomaž Rotovnik, Darinka Verdonik, Zdravko Kačič

Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru  
Smetanova ul. 17, SI-2000 Maribor, Slovenija  
andrej.zgank@uni-mb.si <http://www.elektronika.uni-mb.si>

## Povzetek

V članku bomo predstavili izdelavo govorne baze Broadcast News za slovenski jezik (BNSI). Projekt, ki poteka v sodelovanju med Fakulteta za elektrotehniko, računalništvo in informatiko, Univerze v Mariboru in RTV Slovenija, se je začel leta 2002. Njegov cilj je izdelati govorno bazo, ki jo je možno uporabiti na različnih področjih govornih tehnologij. Najprej bomo predstavili organizacijske vidike izdelave baze ter postopke, ki smo jih pri tem uporabili. Material, ki smo ga vključili v bazo smo zajeli iz arhiva RTV Slovenija. Trenutno poteka obdelava govornega materiala v skladu s specifikacijami, ki so bile določene v okviru projekta. V nadaljevanju članka bomo predstavili prve rezultate razpoznavanja tekočega govora. Sistem smo zasnovali na osnovi akustičnih modelov naučenih na bazi SNABI, ki smo jih prenesli na bazo Broadcast News za slovenski jezik ter ponovno učili. Dosežena je bila 41,7% napaka razpoznavanja besed na stavkih razreda F0.

## Slovenian Broadcast News speech database (BNSI) and LVCSR speech recognition system

This paper presents the Slovenian Broadcast News speech database project that was started in year 2002 as cooperation between Faculty of Electrical Engineering and Computer Science, University of Maribor and Slovenian national broadcaster RTV Slovenija. The goal of the project is to develop a speech database that could be used in different areas of language technologies. First some organizational aspects that were needed in initial phase of the project are described. The raw audio and video material was acquired from the original Analog Beta SP Master tapes that are preserved in the RTV Slovenija's archive. Currently, the speech material is being transcribed using the specifications defined in the framework of the project. In the second part of this paper, first continuous speech recognition results will be presented. The system was based on SNABI acoustic models that were ported on Slovenian Broadcast News database and retrained. Word error rate of 41.7% was achieved on sentences that belong to the F0 focus condition.

## 1. Uvod

Razpoznavanje govora predstavlja eno izmed smeri govornih tehnologij, ki v zadnjem času doživlja prenos iz laboratorijskega okolja v praktično uporabo. Tako je bil v Sloveniji leta 2002 (Žgank et al., 2002) prvič predstavljen javno dostopen govorno voden telefonski portal LentInfo, ki mu je nekaj mesecev kasneje sledil sistem M-Vstopnica (Sket in Imperl, 2002). Značilnost obeh sistemov je, da podpirata razpoznavanje govora za slovenski jezik, vendar samo na nivoju izoliranih besed.

V slovenskem jeziku predstavlja bistveno zahtevnejši problem razpoznavanje tekočega govora z velikim slovarjem besed. Slovenščina sodi v skupino pregibnih jezikov, kar izdatno poveča število vseh možnih besed s katerimi se lahko sreča splošen razpoznavnik tekočega govora. Dodatno težavo predstavlja prosti vrstni red besed v stavku, ki oteži napovedovanje verjetnosti zaporedja dveh ali več besed s pomočjo jezikovnega modela. V slovenskem prostoru je sicer bilo na voljo kar nekaj različnih govornih baz, ki so potrebne za učenje razpoznavnika govora, vendar niso primerne za razpoznavanje tekočega govora z neomejenim naborom besed. Edina delna izjema je baza SNABI (Kačič et al., 2000), katere del je uporaben tudi v te namene, vendar vsebuje samo bran govor.

Upoštevaje ta dejstva smo zasnovali projekt Broadcast News za slovenski jezik (BNSI), katerega cilj je izdelati govorno bazo iz družine Broadcast News, primerno za razpoznavanje tekočega govora z neomejenim naborom

besed. Prva baza iz skupine Broadcast News je nastala leta 1996 na pobudo ameriških organizacij NIST in DARPA (Pallett, 2002). Ideja je, uporabiti pri izdelavi baze "najden" govor, namesto načrtno posnetega. Baze iz družine Broadcast News sestavljajo posnetki različnih televizijskih in radijskih oddaj. Uporabiti jih je možno na različnih področjih govornih tehnologij, kot so na primer: avtomatsko razpoznavanje tekočega govora, indeksiranje govora, TDT, ... Prvi bazi Broadcast News za angleški jezik, ki smo jo uporabljali v preteklosti (Žgank, 1999), so sledile baze v drugih jezikih: španskem, kitajskem, italijanskem, češkem, ... V tujini so bile predstavljene tudi komercialne aplikacije (Brousseau et al., 2003; Imai et al., 2000), zasnovane z bazami Broadcast News. Fakulteta za elektrotehniko, računalništvo in informatiko, Univerze v Mariboru je demonstracijsko aplikacijo, s sistemom za slovensko podnaslavljanje televizijskih oddaj v živo, uspešno predstavila na sejmu TeleInfos, aprila 2004 v Ljubljani.

Začetki projekta Broadcast News za slovenski jezik segajo v leto 2002. Pogodba med partnerjema: Fakulteto za elektrotehniko, računalništvo in informatiko, Univerze v Mariboru in RTV Slovenija, Ljubljana je bila podpisana konec leta 2002. Cilj projekta je izdelati bazo Broadcast News za slovenski jezik, ki jo bo možno uporabiti na različnih področjih govornih tehnologij. Govorna baza bo po dokončanju dostopna širšemu krogu uporabnikov v okviru evropske organizacije ELRA/ELDA. Manjši del baze Broadcast News za slovenski jezik, v obsegu treh

oddaj, je vključen tudi v Panevropsko Broadcast News bazo (COST278-BN domača stran, 2004), ki je nastala v okviru sodelovanja v evropski akciji COST 278 (COST278 domača stran, 2004) in pokriva devet različnih evropskih jezikov.

V članku bomo najprej predstavili nastajanje in izdelavo govorne baze Broadcast News za slovenski jezik. V drugem poglavju bomo opisali priprave in zajemanje materiala. Obdelava govornega materiala bo predstavljena v tretjem poglavju. Predstavitvi govorne baze bo sledila predstavitev razpoznavnika tekočega govora, izdelanega na osnovi baze Broadcast News. V četrtem poglavju bomo podali način tvorjenja akustičnih in jezikovnih modelov. Rezultati razpoznavanja govora se nahajajo v petem poglavju, zaključek in smernice za prihodnje delo pa v zadnjem, šestem, poglavju.

## 2. Priprava in zajemanje materiala

Pri izdelavi baze tipa Broadcast News lahko uporabimo dva različna načina zajemanja materiala. Prvi način je zajemanja materiala neposredno iz etra. V tem primeru vsebuje govorni material tudi vpliv prenosnega kanala, kar lahko olajša izdelavo sistemov, ki bodo neposredno zajemali govor iz etra. Slabost takšnega pristopa je, da je z njim bistveno težje pokriti daljše časovno obdobje, kar je potrebno zato, da res dosežemo neomejen nabor besed. V primeru, ko bi zajemali oddaje samo v krajšem časovnem obdobju, bi lahko prišlo do prevelike medsebojne odvisnosti slovarjev posameznih oddaj, saj pogosto prihaja do ponavljanja tematike znotraj krajšega časovnega obdobja. Drugi možen način zajemanja je snemanje govornega materiala neposredno iz arhiva oddaj. Takšen postopek omogoča bistveno lažje pokritje daljšega časovnega obdobja. V primeru, kadar želimo na takšni bazi zasnovan sistem uporabiti za zajemanje oddaj v živo, si pri odpravljanju vpliva prenosnega kanala pomagamo z metodami robustnega procesiranja govornega signala. Pri izdelavi baze Broadcast News smo uporabili drugi pristop zajemanja govornega materiala, saj je bil naš cilj izdelati govorno bazo za razpoznavanje govora z neomejenim naborom besed. Material za izdelavo baze, ki jo bomo predstavili v nadaljevanju, je bil zajet iz arhiva oddaj RTV Slovenija.

Pred samim začetkom zajemanja govornega materiala smo opravili analizo obstoječih baz iz družine Broadcast News, kjer smo ugotavljali tip in strukturo vključenih oddaj. Na osnovi rezultatov analize in televizijskega sporeda smo pripravili seznam oddaj, ki so bile kasneje vključene v bazo. Pri tem smo kot osnovno obdobje upoštevali čas štirih let – od leta 1999 do leta 2003. Kasnejša analiza je pokazala, da so v televizijskih oddajah govornici moškega spola mnogo pogostejši. Da bi zagotovili enakomerno zastopanost obeh spolov, smo tako zajeli del oddaj tudi iz starejšega obdobja – do leta 1993.

Posnetki oddaj, ki se nahajajo v arhivu, so shranjeni na kasetah Analog Beta SP. Material smo zajemali na dva različna medija. Govorni signal smo shranjevali na DAT Audio kasete, ki smo jih kasneje uporabili kot izhodišče za izdelavo govorne baze. Za pomoč pri obdelavi govornega materiala (identifikacija govorcev, deli s prekrivajočim se govorom), smo na DVD+R nosilce shranjevali slikovni in

govorni material. Podrobnosti o zajetem materialu se nahajajo v tabeli 1.

Tip nosilca	Tip podatkov	Zajemanje
DAT	Zvok	48 kHz vzorčenje 16 bitna kvantizacija 2 kanala
DVD+R	Zvok	48 kHz vzorčenje 16 bitna kvantizacija 2 kanala 256 kbps podatkovni tok AC3 stiskanje
DVD+R	Slika	5.1 Mbps podatkovni tok MPEG2 stiskanje

Tabela 1: Lastnosti zajetega govornega in slikovnega materiala na DAT in DVD+R nosilcih.

V arhivu RTV Slovenija smo tako zajeli 72 ur govornega materiala. Končna verzija baze Broadcast News za slovenski jezik bo vsebovala 36 ur obdelanih posnetkov. Od tega bo 30 ur namenjenih učenju akustičnih modelov, 3 ure razvojnemu prilagajanju sistema in 3 ure testiranju razpoznavnika tekočega govora. V bazo so vključeni sledeči tipi oddaj: TV Dnevnik, Odmevi, Intervju, Polnočni klub, ...

Hkrati z govornim materialom smo iz računalniškega arhiva RTV Slovenija zajeli tudi tekstovni material različnih informativnih oddaj. Pri nastajanju posamezne oddaje novinarji uporabljajo sistem iNews, v katerega vnašajo tekstovni material posameznega prispevka. Med snemanjem oddaje voditelj bere tekst prispevka z monitorja, prav tako se iz tega tekstovnega materiala tvorijo podnapisi oddaje, ki jih RTV oddaja preko teleteksta. Zajeti tekstovni material bomo uporabili za učenje jezikovnih modelov, saj sta oblika in način jezika, ki je v uporabi pri televizijskih oddajah, bistveno drugačna od jezika, ki ga uporabljajo v tiskanih medijih.

## 3. Obdelava govornega materiala

Zajeti govorni material smo v naslednjem koraku obdelali. Za vsako oddajo je potrebno izdelati posebno tekstovno datoteko s prepisom izgovorjenega. Natančno je prepisan vsebovan govor, označeni so drugi zvočni dogodki, ki vplivajo na razpoznavanje govora, označeni so vsi govornici in njihove lastnosti. Pri obdelavi govornega materiala smo uporabljali orodje Transcriber (Barras et al., 2001). Zapisovalcem je pri izdelavi baze zadostovala uporaba osebnega računalnika, saj so bili z njim dostopni vsi potrebni dodatni viri – slika z DVD+R nosilca, Slovenski pravopis, baza časopisa Večer, iskalnika Najdi.si in Google, ...

Pri označevanju govornega materiala smo za osnovo vzeli navodila za izdelavo angleške baze, ki jih je objavila organizacija LDC (LDC domača stran, 2004). Dopolnili smo jih z navodili inštitucije LIMSI (LIMSI domača stran, 2004) in z navodili defi niranimi v okviru iniciative COST 278 Broadcast News.

### 3.1. Nekatera jezikovno odvisna pravila prepisovanja

V nadaljevanju predstavljamo nekatere posebnosti ortografskega prepisovanja baze.

### 3.1.1. Kratice

Kriviculja (~) se uporablja za označevanje kratic, ki so izgovorjene črkovano, npr. ~ZDA, znak @ pa za označevanje kratic, ki so izgovorjene kot ena beseda, npr. @NATO. Morebitne sklanjatve kratic so označene z vezajem, npr. @SAZU-ja, ~APZ-ja, ~BBC-ja, ali s spremenjeno zadnjo črko okrajšave, npr. @NATA.

### 3.1.2. Velike črke

Velike začetnice se pri prepisovanju baze uporabljajo samo v kraticah (~APZ), pri črkovanju (~S ~T ~R ~I ~T ~A ~R) in kot velike začetnice pravih lastnih imen (Moskva, George). Občne besede, ki so del lastnih imen ali lastno ime, so zapisane z malo začetnico in med narekovaji, npr. "ministrstvo za notranje zadeve republike Slovenije", "združene države Amerike".

### 3.1.3. Ločila

Ločila se postavljajo v skladu s slovenskim pravopisom (Toporišič et al., 2001). Dovoljena so naslednja ločila: (.) pika, (,) vejica, (?) vprašaj, (!) klicaj, (:) dvopičje, (...) tri pike. Presledek se vstavi pred ločilom in za njim (dela, dela ...).

### 3.1.4. Okrajšave

Pri prepisovanju je dovoljeno uporabljati tudi nekatere okrajšave: itd., ipd., tj., npr., oz. Okrajšave akademskih nazivov (dr., prof.) niso dovoljene. Tudi števila se vsa izpišejo z besedo (npr. dvajset celih sedem).

## 4. Sistem za razpoznavanje tekočega govora

Razpoznavnik govora, ki temelji na bazi Broadcast News, predstavlja bistveno zahtevnejši sistem kot so tisti sistemi, ki jih uporabljamo v telefonskih aplikacijah ukazovanja z izoliranimi besedami ali v narekovalnih sistemih. Med delovanjem tako naletimo na različne kategorije govora. Primer za razpoznavalnike govora relativno preproste kategorije je bran govor v študijskem okolju. Bistveno zahtevnejša kategorija je na primer spontan govor tujega govorca preko telefonskega kanala slabe kakovosti. Posledica je, da je potrebno v razpoznavnik govora vključiti dodatne module, kot so: modul za ločevanje govora glede na njegove akustične značilnosti, modul za združevanje govorcev v gruče, modul za akustično adaptacijo na različne govorce, modul za prilagoditev aktivnega slovarja trenutnemu besedišču, ... Da bi kar najhitreje omogočili razvoj potrebnih dodatnih modulov, smo s trenutno razpoložljivim govornim materialom iz baze Broadcast News za slovenski jezik izdelali osnovni razpoznavnik govora, ki ga bomo predstavili v nadaljevanju.

### 4.1. Priprava akustičnih modelov

Trenutno obdelani govorni material iz baze Broadcast News ne zadostuje za samostojno učenje kvalitetnih akustičnih modelov. Težavo smo rešili z uporabo postopka prenosa obstoječih akustičnih modelov in njihovim ponovnim učenjem (Žgank et al., 2003). Kot osnovo smo tako vzeli akustične modele, naučene na slovenski študijski bazi SNABI. Izvedli smo standarden postopek

učenja, katerega rezultat so bili notranjebesedni trifonški akustični modeli s 16 kombinacijami Gaussovih porazdelitvenih funkcij verjetnosti na stanje.

Akustične modele naučene na bazi SNABI, je potrebno sedaj prenesti v okolje baze Broadcast News. Pri tem je bilo potrebno prilagoditi nabor fonemov prve – izvorne – baze na nabor fonemov druge – ciljne – baze. Do razlike med obema naboroma je prišlo zaradi postopka mapiranja redkih fonemov, ki smo ga uporabili pri učenju akustičnih modelov z bazo SNABI. Pri prilagoditvi obeh naborov fonemov, smo kot vodilo upoštevali akustično-fonetične lastnosti slovenskih fonemov (Zemljak et al., 2002; Bezljaj, 1960).

Iz učnega korpus baze Broadcast News smo vzeli samo tiste stavke, ki so bili označeni z razredom F0 (LDC domača stran, 2004). V ta razred sodi bran govor v študijskem okolju, brez motenj v ozadju. V naboru posnetkov za ponovno učenje je bilo 349 stavkov, v skupni dolžini 32,3 minut. Enak postopek smo uporabili za testni nabor, ki je vseboval 189 stavkov, v skupni dolžini 16,7 minut.

Prenesene akustične modele smo sedaj uporabili kot izhodišče pri postopku ponovnega učenja modelov z govornim materialom iz baze Broadcast News. V vsakem koraku učenja prikritih modelov Markova smo obnavljali samo srednje vrednosti in uteži kombinacij Gaussovih porazdelitvenih funkcij verjetnosti, s katerimi je predstavljen akustični prostor. Vrednosti varianc in verjetnosti prehodov med posameznimi stanji so ostajale nespremenjene. Akustične modele, ki smo jih dobili kot rezultat postopka ponovnega učenja, smo nato uporabili v razpoznavniku govora.

### 4.2. Priprava jezikovnih modelov

Med razpoznavanjem tekočega govora ima jezikovni model zelo pomembno vlogo, saj modelira zajeti govor na višjem nivoju kot akustični modeli. Kot osnovo za učenje jezikovnega modela smo uporabili besedilni korpus dnevnika Večer, v katerem je zbranih 85 milijonov besed, od tega 750.000 različnih. Slabost uporabe časopisnega besedilnega korpusa je v tem, da je vrst in način besedila, ki ga srečamo v televizijskih oddajah bistveno drugačen od besedila, ki ga zasledimo v časniku.

Da bi kar najbolj zmanjšali to razliko v obliki besedila, smo osnovnemu besedilnemu korpusu dnevnika Večer dodali še del tekstovnega materiala oddaj, ki smo ga zajeli iz sistema iNews. Dodatni besedilni korpus je vseboval 177.000 besed, od tega 26.000 različnih. Na osnovi tega združenega besedilnega korpusa smo izdelali bigramski jezikovni model zasnovan na besedah, ki smo ga uporabili v razpoznavniku govora.

### 4.3. Razpoznavnik tekočega govora z velikim slovarjem besed

Slovar razpoznavalnika govora je vseboval 20.000 različnih besed. Kot osnovo smo vzeli vse tiste besed, ki so nastopale v učnem korpusu – teh je bilo približno 7.000 – ter seznam dopolnili z najpogostejšimi besedami iz besedilnega korpusa Večer. Grafemsko-fonemsko pretvorbo smo v prvem koraku opravili s slovarjem SiFlex. Manjkajoče besede smo pretvorili s pomočjo avtomatske grafemsko-

fonemske pretvorbe. Analiza je pokazala, da slovar razpoznavalnika govora pripravljen na takšen način, ne vsebuje samo 0,01% besed, ki se nahajajo v testnem naboru. Dosežena pokritost je zelo dobra, saj so dosedanje izkušnje pokazale, da lahko za slovenski jezik pri takšni velikosti slovarja pričakujemo približno 80% pokritost (Rotovnik et al., 2003). Za razpoznavanje govora smo uporabili standarden razpoznavalnik govora, ki je deloval v enoprehodnem načinu z bigramskim jezikovnim modelom.

## 5. Rezultati

Rezultate razpoznavanja govora na izbranem testnem naboru stavkov razreda F0 prikazuje tabela 2.

	Rezultat (%)
Pravilne b.	67,9
Zamenjane b.	27,7
Izbrisane b.	4,4
Vrinjene b.	9,6
NRB	41,7

Tabela 2: Rezultati razpoznavanja tekočega govora z bazo Broadcast News.

Predstavljen sistem razpoznavanja tekočega govora je dosegel 41,7% napako razpoznavanja besed (NRB). Največje število napak so doprinesle zamenjane besede (27,7%). Podrobnejša analiza razpoznanih stavkov je pokazala, da lahko delež teh napak pripišemo infleksijski naravi slovenskega jezika, saj je prihajalo do napak v končnicah besed, medtem ko je bila osnova razpoznanana pravilno. V nadaljnjem delu bo pri zasnovi razpoznavalnika govora za bazo Broadcast News za slovenski jezik potrebno uporabiti tudi metode, ki so prirejene za infleksijske jezike.

## 6. Zaključek

V članku smo predstavili govorno bazo Broadcast News za slovenski jezik (BNSI), ki je rezultat projekta med Fakulteto za elektrotehniko, računalništvo in informatiko, Univerze v Mariboru in RTV Slovenija. Baza vsebuje posnetke različnih televizijskih oddaj, predvsem s področja dnevno-informativnega programa. Obdelani govorni material smo v nadaljevanju uporabili za izdelavo razpoznavalnika tekočega govora z velikim slovarjem besed, ki ni bil omejen na določeno besedišče.

Konec projekta izdelave baze Broadcast News za slovenski jezik je predviden za december 2004. Predstavljeni razpoznavalnik govora služi kot osnova za nadaljni razvoj sistema Broadcast News, predvidena je vgradnja dodatnih modulov, ki bodo izboljšali delovanje sistema, še posebej z upoštevanjem posebnosti slovenskega jezika.

## Zahvala

Avtorji članka se zahvaljujo delavcem RTV Slovenija, ki so pomagali izpeljati projekt Broadcast News. Posebna zahvala grem vsem sodelavcem, ki so bili vključeni v zajemanje, transkribiranje in obdelavo baze Broadcast News za slovenski jezik.

## 7. References

- Barras, C., Geoffrois, E., Wu, Z. and Liberman, M., 2001. Transcriber: Development and use of a tool for assisting speech corpora production. *Speech Communication*, Volume 33, Issues 1-2, 5-22.
- Bezljaj, F., 1960. *Osnove fonetike*, Ljubljana.
- Brousseau, J., Beaumont, J.F., Boulianne, G., Cardinal, P., Chapdelaine, C., Comeau, M., Osterrath, F., Ouellet, P., 2003. Automatic Closed-Caption of Live TV Broadcast News in French, *Proc. Eurospeech 2003*, Ženeva, Švica. <http://cost278.org>  
<http://cost278.org/bn>  
<http://www.etca.fr/CTA/gip/Projets/Transcriber/-fr/user.html>  
<http://www ldc.upenn.edu/Projects/Corpus-Cookbook/-transcription/broadcast-speech/english/index.html>
- Imai, T., Kobayashi, A., Sato, S., Tanaka, H., and Ando, A., 2000. Progressive 2-pass decoder for real-time broadcast news captioning. *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, 1937-1940, Carigrad, Turčija.
- Kačič, Z., Horvat, B., Zögling A., 2000. Issues in Design and Collection of Large Telephone Speech Corpus for Slovenian Language. *Proc. Second International Conference on Language Resources and Evaluation*, Atene, Grčija.
- Pallett, D. S., 2002, The role of the National Institute of Standards and Technology in DARPA's Broadcast News continuous speech recognition research program. *Speech Communication*, Volume 37, Issues 1-2, 1:3-14.
- Rotovnik T., Maučec M. S., Horvat B., Kačič Z., 2003. Slovenian Large Vocabulary Speech Recognition with Data-Driven Models of Inflectional Morphology. *IEEE Workshop on Automatic Speech Recognition and Understanding*, Deviški otoki ZDA, 83-88.
- Sket, G., B. Imperl, 2002. M-vstopnica: uporaba avtomatskega razpoznavanja govora v praksi. *Jezikovne tehnologije 2002*, Ljubljana.
- Toporišič, J., et al., 2001. *Slovenski pravopis*. Ljubljana.
- Zemljak, M., Kačič, Z., Dobrišek, S., Gros, J., Weiss, P., 2002. Računalniški simbolni fonetični zapis slovenskega govora. *Slav. rev.*, apr.-jun. 2002, letn. 50, št. 2, 159-169.
- Žgank, A., 1999. *Avtomatsko razpoznavanje tekočega govora neodvisnega govornika* : diplomsko delo univerzitetnega študija, Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru.
- Žgank, A., M. Rojc, B. Kotnik, D. Vljaj, M. Sepesy Maučec, T. Rotovnik, Z. Kačič, A. Zögling Markuš, B. Horvat, 2002. Govorno voden informacijski portal LentInfo - predhodna analiza rezultatov. *Jezikovne tehnologije 2002*, Ljubljana.
- Žgank, A., Kačič, Z., Horvat, B., 2003. Comparison of acoustic adaptation methods in multilingual speech recognition environment. *Text, speech and dialogue : 6th International Conference, TSD 2003*, Česke Budejovice, Češka.