

Inguna SkadiņaSabiedrība *Tilde*Ingona.Skadina@tilde.lv

Jaunas iespējas attēlu meklēšanā : ģeotelpiskajā informācijā un valodu tehnoloģijās balstīta attēlu meklēšanas platforma *TRIPOD*

Kopsavilkums. Biežākie iemesli, kāpēc lasītājs neatrod meklēto, bet datubāzē iekļauto informāciju, ir neprecīzi uzdots vaicājums vai neprecīzi aprakstīts objekts datubāzē. Īpaši aktuāla šī problēma ir attēlu meklēšanā, jo objektiem laika gaitā var mainīties nosaukumi, attēlu paraksti mēdz būt neprecīzi vai pat daudznozīmīgi. Rakstā aprakstīti pētniecības rezultāti un tehnoloģijas, kas ļauj efektīvi meklēt attēlus, kuri saistīti ar konkrētu vietu. Inteliģentā meklēšanas sistēma izmanto zināšanas, kas iegūtas, apvienojot valodu tehnoloģiju, ģeotelpisko informāciju sistēmu un ontoloģiju iespējas. Attēlu parakstu papildināšanai tiek izmantota internetā atrodamā informācija, un tās apstrādē izmantoti daudzdokumentu kopsavilkumu ģenerēšanas rīki. Sistēma ir daudzvalodu, tajā iekļauta angļu, itāļu, latviešu un vācu valoda. Paredzēts, ka pētniecības rezultātus varēs izmantot attēlu meklēšanā, attēlu parakstu papildināšanā un parakstu ģenerēšanā attēliem ar ģeogrāfiskajām koordinātām. Pētījums veikts ES 6. ietvarprogrammas projekta *TRI-Partite multimedia Object Description (Tripod)* ietvaros, izmantojot LNB attēlu kolekcijas.

Ievads

Gan Latvijas Nacionālajā bibliotēkā (LNB), gan citās bibliotēkās daudzu gadu desmitu laikā ir uzkrāts milzīgs attēlu daudzums. Pēdējos gados šīs attēlu kolekcijas tiek digitalizētas, tādējādi tās kļūvušas daudz pieejamākas interesentiem. LNB saviem lasītājiem nodrošina piekļuvi tādām attēlizdevumu kolekcijām kā Latvijas plakātu virtuālā galerija, LNB digitalizēto atklātņu kolekcija, Baltijas Centrālās bibliotēkas digitalizētie attēli u.c. LNB darbs attēlu kolekciju digitalizēšanā turpinās, un tas ļauj lasītājiem arvien pilnīgāk iepazīt Latviju vēsturiskajā aspektā.

Tomēr attēlu kolekciju digitalizēšana neatrisina visas ar attēlu izmantošanu saistītās problēmas – ne mazāk svarīga ir zināšanās balstīta meklēšanas sistēma, kas ļauj atrast attēlus arī gadījumos, ja vaicājums ir neprecīzs, attēla paraksts ir neprecīzs, objektam laika gaitā ir mainīts nosaukums.

Piemēram, lietotājs, ievadot vaicājumu *Jāņa baznīca*, visticamāk vēlas atrast ar Rīgas Sv. Jāņa baznīcu saistītos attēlus, nevis visas baznīcas Latvijā, kuru nosaukumā ir vārds *Jānis*. Savukārt, ja lietotājs vēlas atrast baznīcas Rīgā, tad, ievadot vaicājumu *baznīcas Rīgā*, attēli, kuru parakstos vārds *Rīga* nav ietverts (piem., *Sv. Pētera baznīca* LNB attēlu kolekcijā), netiek atrasti. Visbeidzot daudzu vietu nosaukumi laika gaitā ir mainījušies, piem., *Lībava* (tagadējā *Liepāja*), vai mainījusies to pareizrakstība (piem., *Laudonas basniza*). Iepriekš minētajos gadījumos vienkārša meklēšanas sistēma, kas neizmanto zināšanas, atradīs tikai daļu ar objektu saistīto attēlu, kamēr citi attēli paliks neatrasti, t.i., meklēšanas sistēmai būs zema precizitāte.

Raksta mērķis ir iepazīstināt ar iespējām pilnveidot ar vietām saistīto attēlu meklēšanu, izmantojot valodu tehnoloģijas, ģeotelpisko informāciju un ontoloģijās apkopoto informāciju. Tādējādi dators ne tikai atradīs lietotāja izvēlētos attēlus pēc atslēgvārdiem, bet arī varēs veikt daudz sarežģītākus uzdevumus:











- izmantojot ģeotelpisko informāciju, ontoloģijas un valodu tehnoloģijas, atrast attēlus, kuros attēlotā vieta norādīta neprecīzi vai laika gaitā tās nosaukums mainījies;
- izmantojot interneta resursus, kopsavilkumu ģenerēšanas rīkus un ģeotelpisko informāciju, papildināt vai izveidot parakstus attēliem, kuriem paraksti ir neprecīzi vai to nav vispār;
- izmantojot valodu tehnoloģijas, ģenerēt attēlu parakstus lietotājam vēlamajā valodā arī gadījumos, kad attēla sākotnējais paraksts ir kādā citā valodā.

Raksta 1. nodaļā aprakstītas attēlu meklēšanas iespējas, kombinējot ontoloģijas, ģeogrāfisko informāciju un valodas tehnoloģijas; 2. nodaļā aprakstītas tehnoloģijas, kas ļauj veidot attēla parakstus attēliem ar ģeotelpisko informāciju, 3. nodaļā aprakstītas daudzvalodu daudzdokumentu kopsavilkumu ģenerēšanas tehnoloģijas. Darbā izmantotas LNB attēlu kolekcijas ar vairāk nekā 4000 attēlu. Pētījums veikts ES 6. ietvarprogrammas projekta *TRI-Partite multimedia Object Description (Tripod)* ietvaros (<http://tripod.shef.ac.uk/>; I. Skadiņa un A.Vasiljevs, 2008), pētījuma rezultāti tiek implementēti *Tripod* attēlu meklētājā (skat. 1. attēlu).

Search Results

add selected to lightbox [edit metadata](#) [Download XMP file](#) 46 images found

select [all/none](#) 1 2 3 4 5 next [print](#) images per page 10

				
<input type="checkbox"/> Riga Castle, 15th-17t...	<input type="checkbox"/> Old Riga	<input type="checkbox"/> Riga - Alexander-Boul...	<input type="checkbox"/> Riga - Alexander Boulv.	<input type="checkbox"/> Latvian SSR, Riga, Sq...
				
<input type="checkbox"/> Latvian SSR, Riga, Sq...	<input type="checkbox"/> Latvian SSR, Riga, St...	<input type="checkbox"/> Riga, Polytechnical I...	<input type="checkbox"/> Riga, Polytechnical I...	<input type="checkbox"/> Riga, View of the Pol...

1. attēls. LNB attēlu kolekcijas fragments *Tripod* meklētājā, meklējot pēc atslēgvārda 'Riga'

1. ATTĒLU MEKLĒŠANA, IZMANTOJOT ĢEOTELPISKO INFORMĀCIJU, ONTOLOĢIJAS UN VALODU TEHNOLOĢIJAS

Lai attēlus varētu sameklēt gadījumos, kad attēla paraksts nesakrīt ar meklēšanas kritērijā uzdotajiem vārdiem, projektā *Tripod* ir izstrādātas tehnoloģijas, kas attēla metadatos pievieno tā aprakstošos jēdzienus, kuri iegūti, izmantojot toponīmu un jēdzienu ontoloģijas, valodas tehnoloģijas un ģeotelpisko informāciju. Tādējādi *Tripod* meklētājs ļauj atrast attēlus pēc atslēgvārdiem gan gadījumos, kad atslēgvārds ir iekļauts attēla parakstā, gan metadatos esošajā informācijā.

Ar attēlu saistīto jēdzienu izguve notiek vairākos secīgos soļos (skat. 2. attēlu).

Wp3: Latvian

[Change parameters](#)

Workflow Progress	
Completed	✓ Toponym Ontology Service - getGeoTaggedCaption
	✓ Concept Ontology Service - getConceptsForToponym
	✓ Concept Ontology Service - describeConcept
	✓ Web Keyword Confidence Service - getConceptKeywordsWithConfidence
	✓ Concept Ontology Service - describeConcept
	✓ Web Keyword Confidence Service - getConceptKeywordsWithConfidence

2. attēls. Jēdzienu izguves procesa no attēla paraksta vispārīgā shēma

Lai attēlam piekārtotu saistītos jēdzienus, vispirms, izmantojot toponīmu ontoloģijas servisu (*Toponym Ontology Service*), tiek veikta attēla paraksta analīze, atrasti toponīmi un tiem piekārtota atrašanās vieta kartē. Pēc tam, izmantojot jēdzienu ontoloģijas servisu (*Concept Ontology Service*), atrastajiem toponīmiem tiek piekārtoti iespējamie saistītie jēdzieni (metode *getConceptsforToponym*). Lai atrastu iespējami saistītos jēdzienus, tiek izmantoti dati no ģeotelpisko datu datubāzes (pašlaik *Corine*), kas tiek sasaistīti ar vietu tipiem (piemēram, pilsēta, ezers, kalns u.c.) no jēdzienu ontoloģijas (metode *describeConcept*).

Nākamajā solī papildu informācija par toponīmu un ar to saistītiem jēdzieniem tiek meklēta internetā un tiek novērtēta atbilstība (metode *getConceptKeywordsWithConfidence*). Visbeidzot katra jēdziena atbilstība tiek pārvērtēta, pamatojoties uz interneta meklēšanas rezultātiem, izgūto ģeogrāfisko informāciju un jēdzienu ontoloģijas statistiku (metode *getConceptKeywordsWithConfidence*). Izvēlētie jēdzieni pēc tam tiek sakārtoti un filtrēti pēc atbilstības pakāpes.

TOPONĪMU ONTOLOĢIJAS SERVISS

Toponīmu ontoloģija ir viens no centrālajiem komponentiem *Tripod* sistēmā. Ontoloģija ietver hierarhiski apkopotu informāciju par vietām kā jēdzieniem un to īpašībām. Lai varētu izmantot toponīmu ontoloģijas datus, vispirms jāveic attēla paraksta analīze, noskaidrojot tajā ietvertos vietvārdus. Latviešu valodā šim mērķim izmanto automātisko teksta marķēšanas programmatūru (*tagger*), kas ietver latviešu valodas morfoloģiskās analīzes rīku un, izvērtējot kontekstu, katram vārdam piekārt visvarbūtiskāko gramatisko raksturojumu, t.i., informāciju par vārdšķiru un vārda formas raksturojumu. Piemēram, analizējot attēla parakstu *Daugavas klinšu krasts*, XML formātā tiek iegūta šāda informācija:

```

    <TWSTaggedCaption>
  <POSElement>
    <TWSPoSWord>Daugavas</TWSPoSWord>
    <TWSType>NNP</TWSType>
    <TWBasform>Daugava</TWBasform>
  </POSElement>
  <POSElement>
    <TWSPoSWord>klinšu</TWSPoSWord>
    <TWSType>NNS</TWSType>
    <TWBasform>klints</TWBasform>
  </POSElement>
  <POSElement>
    <TWSPoSWord>krasts</TWSPoSWord>
    <TWSType>NN</TWSType>
    <TWBasform>krasts</TWBasform>
  </POSElement>
</TWSTaggedCaption>

```

Informācija par katru vārdu aprakstīta tagos <POSElement>, kas savukārt ietver vārda formu, kāda tā ir tekstā (tags <TWSPoSWord>), gramatisko informāciju (tags <TWSType>, NNP – īpašvārds vienskaitlī, NNS – lietvārds daudzskaitlī, NN – lietvārds vienskaitlī) un vārda pamatformu (tags <TWBasform>).

Pēc tam, kad veikta attēla paraksta analīze, notiek toponīmu atpazīšana. Šim mērķim izmantotas divas metodes:

- 1) salīdzināšana ar ģeogrāfisko nosaukumu rādītāju,
- 2) nosaukumu atpazīšana pēc šablona.

Salīdzināšanai ar ģeogrāfisko nosaukumu rādītāju projektā *Tripod* tiek izmantoti Latvijas Republikas Valsts zemes dienesta sagatavotie „Latvijas Republikas ģeogrāfiskie nosaukumi”. Tie ietver Latvijas nozīmīgāko ģeogrāfisko objektu nosaukumus – pilsētu, ciemu, galveno Baltijas jūras un tās piekrastes objektu, lielāko ezeru, garāko upju, augstieņu un to augstāko kalnu nosaukumus. Izmantojot šo avotu, piemēram, attēla parakstā *Rušonu ezers pie Jaunaglonas*, tiek atrasti divi toponīmi (tagos <toponym>) un to koordinātas (tagos <lattitude> un <longitude>):

```

- <toponyms>
- <item>
  <toponym>Jaunaglona</toponym>
  <type>village</type>
  <lattitude>56°10'</lattitude>
  <longitude>27°00'</longitude>
</item>
- <item>

```

```

<toponym>Rušona ezers</toponym>
<type>lake</type>
<latitude>56°11'</latitude>
<longitude>27°02'</longitude>
</item>
</toponyms>

```

Izvērtējot pirmās metodes piemērotību uzdevumam, nācās secināt, ka izvēlētā avota pārklājums ir nepietiekams: tajā nav ietverti daudzi ūdeņu nosaukumi un mazākas apdzīvotās vietas, tāpat tas neietver tādus objektus kā, piemēram, ielas, parkus u.c. Tāpēc sākotnējā metode tika papildināta ar toponīmu atpazīšanu pēc šablona, kas gan neļauj toponīmam piesaistīt koordinātas. Piemēram, attēla parakstā *Rīga. Brīvības piemineklis ziemā* tiek atpazīti divi toponīmi – *Brīvības piemineklis* un *Rīga* –, no kuriem tikai vienam ir iespēja noteikt koordinātas:

```

= <toponyms>
= <item>
  <toponym>Brīvības piemineklis</toponym>
  </item>
= <item>
  <toponym>Rīga</toponym>
  <type>city, town</type>
  <latitude>56°57'</latitude>
  <longitude>24°06'</longitude>
  </item>
</toponyms>

```

Lai izvērtētu izvēlētās metodes lietderību, toponīmu izgūšanas metodes precizitāte un pārklājums (*recall*) tika novērtēts testa korpusam ar 100 patvaļīgi izvēlētiem attēlu parakstiem no LNB attēlu kolekcijas un no portāla *letonika.lv*.

2. ATTĒLU PARAKSTU ĢENERĒŠANA ATTĒLIEM AR ĢEOGRĀFISKĀM KOORDINĀTĀM

Ar jaunāko tehnoloģiju palīdzību var vienkārši iegūt attēlus, kuriem jau ir pievienotas ģeogrāfiskās koordinātas vai pat pilnīgāks ģeotelpiskais raksturojums, t.sk. augstums un leņķis. Šādas iespējas ietvertas gan vairākos mobilajos tālruņos, piem., *Nokia E95*, *HTC G1*, gan fotokamerās, piemēram, *Ricoh 500SE*.

Ja attēliem pievienota ģeogrāfiskā informācija, tad šādu attēlu parakstus var ģenerēt automātiski, sasaistot ģeogrāfisko informāciju ar ontoloģijās, datubāzēs un internetā atrodamo informāciju. Tomēr šādu attēlu parakstu ģenerēšana ir sarežģīts uzdevums, līdzīgi kā meklēšanā, arī attēlu parakstu un aprakstu veidošanā tiek izmantotas valodu tehnoloģijas, ģeogrāfiskā informācija un ontoloģijas.

Attēlu parakstu ģenerēšana sākas ar attēlam pievienotās ģeogrāfiskās informācijas analīzi, kuras mērķis ir noskaidrot gan attēla uzņemšanas vietu, gan ar to saistīto ģeogrāfisko

informāciju, piemēram, reljefu, ģeometriju, objekta tipu u.c. Izmantojot šo informāciju, iespējams ģenerēt attēla paraksta apraksta veidni XML formātā, no kuras pēc tam varēs ģenerēt attēlu parakstus četrās valodās: angļu, itāļu, latviešu un vācu. Piemēram, lai ģenerētu parakstu *fotogrāfija uzņemta pie Brīvības pieminekļa Rīgā*, nepieciešama šāda XML veidne:

```
<machine-readable>
<caption-element caption-element-type="sentence">
  <caption-element caption-element-type="photo-taken"/>
  <caption-element caption-element-type="proximal">
    <caption-element caption-element-type="toponym"
      value="Freedom Monument"/>
  </caption-element>
  <caption-element caption-element-type="contained-in">
    <caption-element caption-element-type="toponym" value="Riga"/>
  </caption-element>
</caption-element>
</machine-readable>
```

Ar XML tagu <caption-element> palīdzību tiek aprakstīta gan attēla uzņemšanas vieta (atribūts `caption-element-type="proximal"`, ietverot atribūtu `caption-element-type="toponym"` kopā ar atribūtu `value`), gan objekta atrašanās vieta (atribūts `caption-element-type="contained-in"`, ietverot `caption-element-type="toponym"` kopā ar atribūtu `value`).

Attēlu parakstu ģenerēšanai latviešu valodā izstrādāts toponīmu transliterēšanas serviss (T. Gornostay un I. Skadiņa, 2009), kas toponīmus angļu valodā, kuru tulkojumi nav fiksēti vārdnīcās, atveido („tulko”) atbilstoši īpašvārdu atveides principiem latviešu valodā (A. Ahero, 2006).

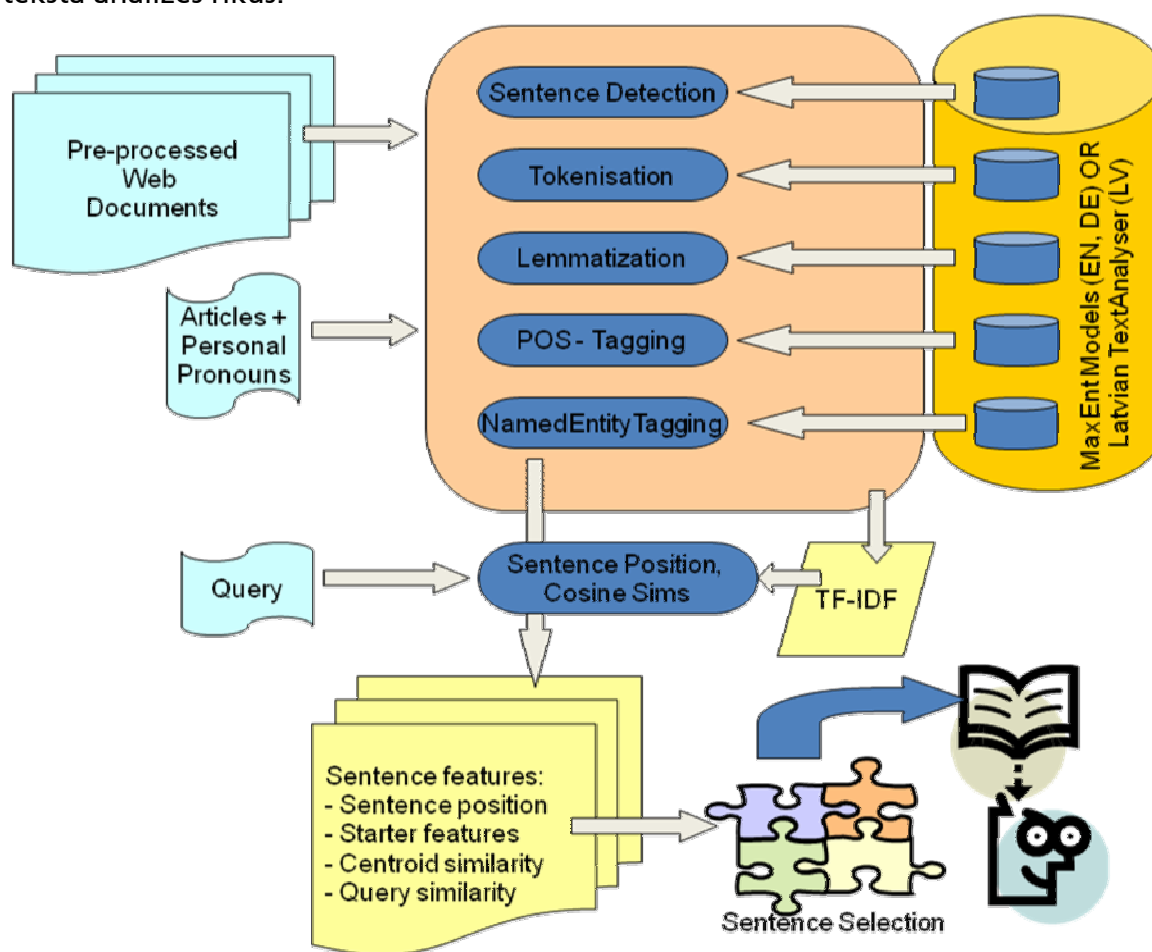
Sistēmā ietverti 110 šablonu angļu toponīmu transliterēšanai, kas ļauj transliterēt gan atsevišķus vārdus, gan vārdu savienojumus. Kaut gan īpašvārdu atveides pamatprincipi latviešu valodā ir definēti, tie nav viennozīmīgi. Tāpēc šablonu izveidei tika analizēti vairāki fūkstošī angļu valodas toponīmu no datubāzes *Geonames* (www.geonames.org). Tā kā toponīma tulkojums var atšķirties dažādiem toponīmu tipiem (piemēram, *Bexhill* tulko kā *Bekshila*, ja tā ir apdzīvota vieta, bet kā *Bekshils*, ja tas ir kalns), tad, izstrādājot algoritmu, tika analizēti 59 biežāk lietotie toponīmu tipi.

Izstrādātais algoritms ļauj transliterēt toponīmus, kurus veido viens vārds, un izvēlēties transliterēšanas vai tulkošanas stratēģiju vārdu savienojumos. Piemēram, vārdu savienojums *Green Isle* tiek transliterēts kā *Grīnaila*; savukārt vārdu savienojums *Newton Point* tiek daļēji tulkots un daļēji transliterēts kā *Ņūtona zemesrags*.

3. KOPSAVILKUMU ĢENERĒŠANA, IZMANTOJOT INTERNĒTA RESURSUS

Līdztekus attēlu parakstiem, attēliem ar ģeotelpisko informāciju plānots pievienot plašāku informāciju par attēlā redzamo objektu. Lai šādu informāciju iegūtu, tiek izmantota kopsavilkumu veidošana no daudziem dokumentiem (skat. 3. attēlu). Dokumentu atrašanai

tiek izmantoti tradicionālie informācijas meklēšanas rīki, piemēram, *Google* un *Yahoo*. Kopsavilkumu ģenerēšanai tiek izmantoti pirmie piecpadsmit meklētāja atrastie dokumenti, kurus analizējot tiek veidoti līdz 200 vārdu garī kopsavilkumi (A. Aker un R. Gaizauskas, 2008). Tehnoloģija ir universāla, tomēr ietver no valodas atkarīgus komponentus, piemēram, teksta analīzes rīkus.



3. attēls. Daudzu dokumentu kopsavilkumu ģenerēšanas rīka vispārējā darbības shēma

Kopsavilkuma veidošana notiek trīs soļos:

- dokumentu lingvistiskā analīze,
- nozīmīgu teikumu noteikšana,
- teikumu izvēle.

Teksta analīze sākas ar teikuma robežu noteikšanu (*Sentence detection*), t.i., teksts vispirms tiek sadalīts teikumos un turpmākās metodes strādā ar katru teikumu atsevišķi. Kad teikuma robežas ir noteiktas, tas tiek sadalīts vienībās (*tokenisation*) – vārdos, pieturzīmēs, saīsinājumos, skaitļos u.c. Pēc tam teikumā esošie vārdi tiek analizēti (*POS – Tagging*) un tiek

atrasta to pamatforma (*Lemmatization*). Valodām, kurām ir nosaukumu (*named entity*) noteikšanas rīki, tiek veikta nosaukumu atpazīšana tekstā.

Kad ieejas dokumentu lingvistiskā analīze ir veikta, tiek analizēts katra teikuma nozīmīgums (svars) un skaitliski novērtētas dažādas teikumam piemītošās īpašības:

- teikuma vieta dokumentā,
- teikuma sākuma īpašības,
- līdzība ar centroīdu,
- līdzība vaicājumam,
- līdzība modelim.

Teikuma izvēles solī secīgi tiek izvēlēti teikumi ar lielāku nozīmību tā, lai kopsavilkums nepārsniegtu 200 vārdus.

Metode uzrāda labus rezultātus gadījumos, kad internetā ir nepieciešamais daudzums ar izvēlēto tēmu saistītu dokumentu (skat. 4. attēlu). Gadījumos, kad atbilstošo dokumentu ir maz, metode ir kļūdaina, jo teikumi kopsavilkumam var tikt izvēlēti no dokumentiem, kas tikai daļēji atbilst vaicājuma nosacījumiem.

Piemineklis personificē Latvijas brīvību un neatkarību. Par Brīvības pieminekļa celšanu sāka runāt pēc atbrīvošanas cīņu izbeigšanās, bet tieši šai lietai kā pirmais piegāja ministru prezidents Z.Brīvības piemineklis atrodas Latvijas galvaspilsētas Rīgas centrā, Brīvības bulvārī, pie Vecrīgas. Latviešu tautas brīvības simbols - Brīvības piemineklis, uzcelts par tautas saziēdotiem līdzekļiem un svinīgi atklāts 1935. gada 18. novembrī. Pieminekļa izveidē labi akcentējas gan tektonika, gan tēlniecība, kur uzsvars likts uz simboliski alegorisku tās traktējumu. Trīs zeltītās zvaigznes izkaluši metālkalēji J.Zobens un A.Naika. Padomju okupācijas laikā ziedu nolikšana un pulcēšanās pie pieminekļa tika aizliegta. Minētā gada 11. jūlija "Latvijas Kareivja" numurā lasām šādu ziņojumu: "Piemineklis kritušiem Latvijas atbrīvotājiem. Virs 19 m augstā obeliska slejas 9 m augsta varā veidota brīvības statuja (Milda), kas ir zviedru tēlnieka Ragnara Mirsmēdena darbs. Simboliskā Latvijas Brīvības tēla pakājē (kopējais augstums 42 m) izvietojas skulpturālas grupas – "Latvijas", "Lāčplēsis", "Vaidelotis", "Važu rāvējs". Izstieptajās rokās tā tur 3 piecstūrainas zeltītas zvaigznes, kas simbolizē Latvijas novadus – Vidzemi, Kurzemi un Latgali.

4. attēls. Automātiski ģenerēts kopsavilkums no internetā atrodamajiem dokumentiem par Brīvības pieminekli

KOPSAVILKUMS UN SECINĀJUMI

Rakstā sniegts ieskats jaunākajās daudzvalodu attēlu meklēšanas tehnoloģijās, kas tiek izstrādātas ES 6. ietvarprogrammas projekta *TRI-Partite multimedia Object Description (Tripod)* ietvaros. Pētniecības rezultāti ļaus uzlabot ar vietām saistīto attēlu meklēšanu, kā arī automātiski ģenerēt attēlu parakstus un aprakstus, gadījumos, ja attēliem ir ģeotelpiskā informācija.

Pašlaik rit projekta pēdējais gads, tāpēc iespējams izdarīt secinājumus par pirmajiem iegūtajiem rezultātiem. Projektā sasniegtais ļauj secināt, ka attēlu parakstu automātiska papildināšana un daudznozīmības novēršana ir sarežģīts un zināšanu ietilpīgs uzdevums. Kā galvenās grūtības minamas parakstu lakonisms un informācijas neprecizitāte. Kaut arī iegūto rezultātu izvērtēšana ir nākamā pusgada uzdevums, pašlaik iegūtie rezultāti ir daudzsolīši – izstrādātās tehnoloģijas novērš daudznozīmību un ļauj attēliem piesaistīt tos raksturojošos jēdzienus. Tāpat pirmie eksperimenti ar daudzdokumentu kopsavilkumu ģenerēšanu un attēlu parakstu ģenerēšanu uzrādījuši labus rezultātus.

Tomēr jau tagad iezīmējas vairākas problēmas, no kurām galvenā ir nepieciešamās informācijas trūkums, kas liedz pilnvērtīgi izmantot tehnoloģijas. Tādējādi pašlaik tehnoloģijas lietojamas labi zināmu objektu aprakstīšanai, bet mazāk zināmiem objektiem tās var nedot gaidītos rezultātus.

Izmantotā literatūra

Ahero, Antonija. *Angļu īpašvārdu atveide latviešu valodā*. Rīga: Zinātne, 2006.

Aker, Ahmet; Gaizauskas, Robert. Evaluating automatically generated user-focused multi-document summaries for geo-referenced images. No: *The 22nd International Conference on Computational Linguistics - Coling 2008*. Manchester: 2008.

Gornostay, Tatiana; Skadiņa, Inguna. English-Latvian Toponym Processing: Translation Strategies and Linguistic Patterns. No: *Proceeding of the 13th Annual Meeting of the European Association for Machine Translation EAMT 2009*. Barselona, 2009.

Skadiņa, Inguna; Vasiljevs, Andrejs. Jaunākie sasniegumi latviešu datorlingvistikas lietojumprogrammās. No: *Letonikas otrais kongress. Valodniecības raksti-2*. Rīga: LZA, 2008. 118.-127. lpp.

Summary. Currently searching from photographic libraries offers a quality of retrieval similar to that provided by text document search systems of the 1980s. Among the reasons why retrieval of existing information fails are vague query and inaccurate information in data base. It is especially true for image retrieval, because image captions are almost never updated, despite the fact that the significance of items in an image changes over time. The paper describes research results and techniques for efficient retrieval of images related to location (e.g., landscapes, buildings, etc.). The image retrieval engine uses information obtained through the synergy of geographic information science, information retrieval and

multi-document summarisation in conjunction with Content Based Image Retrieval. The proposed techniques are multilingual, they have been researched for English, German, Italian and Latvian. A prototype system will be produced that provides an integrated automated caption creation system along with a fully functioning image search engine that provides spatial search over the collections of images. The research is supported by Sixth Framework project *TRI-Partite multimedia Object Description (Tripod)*, for caption augmentation task image collection of National Library of Latvia was used.