

The Corpus of the Saeima, Parliament of Latvia

Roberts Dargis

Ilze Auziņa



Institute of Mathematics and Computer Science
University of Latvia

Structure of the parliament

- Saeima – The 100-seat unicameral Latvian parliament;
- Elected by direct popular vote every four years;
- The work of the Saeima is supervised by Presidium;
- Presidium consists of the Speaker, two Deputy Speakers, the Secretary and the Deputy Secretary;

History of Latvia

- 1918: the Republic of Latvia was proclaimed
- 1922: the 1st Saeima - the 4th Saeima
- 1934: The 4th Saeima was dissolved after the coup, Saeima's functions were taken over by the Cabinet of Ministers
- 1940: the Supreme Council of the Soviet Latvia
- 1990: the Supreme Council of the Republic of Latvia
- 1993: Election of the 5th Saeima

Data published by the government

- Transcripts from 1st to 4th term are available in OCR;
- Transcripts since 5th term (year 1993) are available in digital format (html);
- Since 7th term (year 1998) bills are available and annotated. Vote results are available;
- Since 8th term (year 2002) audio recordings are available;
- Since 9th term (year 2006) video recordings are available

Processed data

- 24 years (from 1993 to nowadays);
- 8 parliamentary terms (from 5th to 12th);
- 647 speakers grouped in 7 categories and 83 subcategories;
- 506k speeches containing about 26M words.

- New data is being added regularly.

Categories of the speakers

- “Members of parliament” further divided into parliamentary groups;
- “Presidium”;
- “Representatives of Institutions of Latvia”;
- “Representatives of Ministries”;
- “Government” – ministers and presidents;
- “Foreign Visitors”, for example, foreign presidents, representatives of foreign parliaments, EU, NATO.

Available formats: NoSketch Engine

- Contains speeches from the members of the parliament;
- Automatically morphologically annotated and lemmatized;
- Can be filtered by:
 - Name,
 - Gender, age at the time of speaking;
 - Political party;
 - Time period.

<http://nosketch.korpuss.lv>

Available formats: ParliSearch

- Internally developed system for the corpus of the Saeima
- Main features:
 - full text stem search;
 - ordering by relevance or date;
 - result filtering by period of time;
 - searching for specific speakers or positions they represent;
 - statistics.
- Available corpora
 - saeima.kospuss.lv – the Corpus of the Saeima
 - europarl.korpuss.lv – the Corpus of the EuroParl

Search interface

Teksts



No

Līdz

Sēdes tips

kārtējā noslēguma svinīgā ārkārtas ārkārtas sesijas

- Deputāti
- Latvijas institūciju pārstāvji
- Ministriju amatpersonas
 - Ministru prezidenta biedri
 - Ministru prezidenta kandidāti
 - Ministru kabineta pārstāvji
 - Anatolijs Gorbunovs (1942)
 - Artūrs Kodoliņš
 - Artūrs Kodoliņš
 - Edvīns Inkēns (1958)
 - Guntis Grūbe
 - Ilma Čepāne (1947)
 - Ilmārs Rimšēvičs
 - Ivars Godmanis (1951)

Kārtot pēc

Periods

- ✓ ...
- Saeimas
- 5. Saeima
- 6. Saeima
- 7. Saeima
- 8. Saeima
- 9. Saeima
- 10. Saeima
- 11. Saeima
- 12. Saeima
- Valdības
- 1993/03/08 Valdis Birkavš
- 1994/09/19 Māris Gailis
- 1995/12/21 Andris Šķēle
- 1997/02/13 Andris Šķēle
- 1997/08/07 Guntars Krasts
- 1998/11/26 Vilis Krištopans**
- 1999/07/16 Andris Šķēle
- 2000/05/05 Andris Bērziņš
- 2002/11/07 Einars Repše
- 2004/03/09 Indulis Emsis
- 2004/12/02 Aigars Kalvītis
- 2006/11/07 Aigars Kalvītis
- 2007/12/20 Ivars Godmanis
- 2009/03/12 Valdis Dombrovskis
- 2010/11/02 Valdis Dombrovskis
- 2011/10/24 Valdis Dombrovskis
- 2014/01/22 Laimdota Straujuma
- 2014/11/05 Laimdota Straujuma
- 2014/11/05 Māris Kučinskis

Result for “medicīna atalgojums” “medicine wage”

150 latu, 2007. gada tas būs 462 latu. Mēs varam apskatīt arī arī kādu citu jomu - arhivu jomu ...
Ļoti būtiski ir arī tas, ka mēs rūpējamies par kultūras pieminekļu aizsardzību. Kā ir ar šo inspektoru **atalgojumu**? 2004. gadā, sastādot budžetu, tam paredzēja tikai 194 latus, nākamgad tie būs 656 latu

.....

... . Jau gandrīz gadu ir aktuāla ieslodzījuma vietu **medicīnas** darbinieku **atalgojuma** problēma. Pašreizējā jaunākā un vidējā medicīniskā personāla darba alga ir 30 līdz 35 latu, ārstu alga ir 40 līdz



Jānis Straume (1962)
Saeimas priekšsēdētāji
2001/02/22 kārtējā sēde

... . Uzklaušīsim Latvijas Republikas Ministru prezidenta Andra Bērziņa atbildi uz jautājumu “Par **medicīnas** māsu **atalgojumu** Latvijā”, ko uzdevuši deputāti Bojārs, Čevers, Salkazanovs, Zvejsalnieks un Burvis

... Cienījamie kolēģi! Vienā no Sabiedrības veselības apakškomisijas sēdēm veselības ministrs apgalvoja, ka tikai 39 miljoni ir nepieciešami, lai palielinātu **medicīnas** darbiniekiem **atalgojumu**

Medicīnas darbinieku atalgojuma problēma. 2004. gadā, sastādot budžetu, tam paredzēja tikai 194 latus, nākamgad tie būs 656 latu

Statistics

Statistika

Atrasti 40 708 no 506 352 (8.04%) fragmentiem laika posmā no līdz .

Sadalījums pēc dzimuma

◆ Dzimums	◆ Skaits	◆ Grupas lielums	◆ Procenti no rezultātu skaita ⓘ	◆ Procenti no kategorijas kopskaita ⓘ
Vīrieši	20 593	368 997	50.59%	5.58%
Sievietes	20 115	137 355	49.41%	14.64%

Sadalījums pēc grupas

◆ Grupa	◆ Skaits	◆ Grupas lielums	◆ Procenti no rezultātu skaita ⓘ	◆ Procenti no kategorijas kopskaita ⓘ
Deputāti/Latvijas Reģionu apvienība	1 255	1 255	3.08%	100.00%
Deputāti/No sirds Latvijai	1 080	1 080	2.65%	100.00%
Deputāti/Pie frakcijām nepiederošs	127	7 823	0.31%	1.62%
Deputāti/Reformu partija	5	7 064	0.01%	0.07%
Deputāti/Saskaņa	2 057	2 057	5.05%	100.00%
Deputāti/Saskaņas centrs	71	6 220	0.17%	1.14%
Deputāti/VL-TB/LNNK	3 266	7 819	8.02%	41.77%

Publications

- DARGIS, Roberts, et al. "**ParliSearch—A System for Large Text Corpus Discourse Analysis.**" *Human Language Technologies—The Baltic Perspective: Proceedings of the Seventh International Conference Baltic HLT 2016*. Vol. 289. IOS Press, 2016.



Available formats: processed, machine readable data

- Data dump is not published, but available if requested

Attribution-ShareAlike 4.0 International



Thank you!

Questions?