

SIW schedule

№	SIW	Types of SIW performance	Times for SIW (study week)
1	Text segmentation techniques	Writing	Week 5
2	Latent Dirichlet Allocation (LDA) uses for text segmentation	Writing	Week 9
3	Create a UML model for LDA text segmentation	Writing	Week 14

On the subject "Intellectual Modeling" Tasks and guidelines for SIW

№ 1 SIW

Topic: Text segmentation techniques.

The purpose of the assignment: To be able to apply the knowledge in practice acquired by the student during the lesson.

Submission form: Writing and oral

Assignments: According to given the topic, write a small size of essay.

Methodical instructions: Proper use of Internet resources.

References:

1. Visweswariah, P.D, Wiratunga, K., Sani N.S. (2012). Two-part segmentation of text documents. In: Proceedings 21st ACM International Conference on Information Knowledge Management—CIKM'12 (p 793). ACM, New York: Maui.
2. Scaiano, M., Inkpen, D., Laganière, R., & Reinhartz, A. (2010). Automatic text segmentation for movie subtitles. In: Lecturer Notes Computer Science (pp. 295–298). Springer.
3. Oh, H., Myaeng, S. H., & Jang, M.-G. (2007). Semantic passage segmentation based on sentence topics for question answering. Information Science (Ny), 177, 3696–3717.
4. Song, F., Darling, W. M., Duric, A., & Kroon, F. W. (2011). An iterative approach to text segmentation. In: 33rd European Conference on IR Resources ECIR 2011, Dublin (pp. 629– 640). Berlin, Heidelberg: Springer.
5. Oyedotun, O. K., & Khashman, A. (2016). Document segmentation using textural features summarization and feedforward neural network. Applied Intelligence, 45, 1–15.
6. Flejter, D., Wieloch, K., & Abramowicz, W. (2007). Unsupervised methods of topical text segmentation for polish. SIGIR'13 (pp. 51–58). Dublin: ACM.

YOUR ESSAY SHOULD HAVE THE FOLLOWING STRUCTURE:

- Introduction: Introduce the subject and state your thesis. The objective is to demonstrate that you understand the essay contest question and have formed a response to it.
- Introduction: Introduce the subject and state your thesis. The objective is to demonstrate that you understand the essay contest question and have formed a response to it.
- Conclusion: Summarize the research and analysis presented in the essay and set forth your recommendations and conclusions. Drawing on ideas already presented, you should demonstrate that your conclusions support the thesis you put forward in the opening

paragraphs. Your aim is to convince the reader that your thesis, facts, and analysis are reasonable and valid.

Essays that use a variety of sources—academic journals, news magazines, newspapers, books, government documents, publications from research organizations—fare better in the contest. Web sites and Web pages should not be the only source of information for your essay. When citing Internet sources, include the following information: author(s), title of work, Internet address, and date information was accessed.

YOUR ESSAY MUST:

- address all parts of this year's contest question in English;
- NOT have your name, teacher's name, or school name anywhere on the essay;
- Font must be no smaller than 12 point Times or similarly sized font;
- have a title and page numbers;
- have no more than 1,500 words. The word count limit includes all words in the text but does not include the bibliography and notes (you may exclude the words in the in-text citation from your word count);

№ 2 SIW

Topic: Latent Dirichlet Allocation (LDA) uses for text segmentation

The purpose of the assignment: To be able to apply the knowledge in practice acquired by the student during the lesson.

Submission form: Writing and oral

Assignments: Based on the SIW1, to add a technology of LDA to expand your first contents of essay.

Methodical instructions: Proper use of Internet resources.

References:

1. Brants, T., Chen, F., & Tsochantaridis, I. (2002). Topic-based document segmentation with probabilistic latent semantic analysis. CIKM'02 (pp. 211–218). Virginia: ACM.
2. Doug Beeferman, Adam Berger, and John Lafferty. 1999. Statistical models for text segmentation. Machine learning, 34(1):177–210.
3. David M. Blei, Andrew Y Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. Journal of Machine Learning Research, 3:993–1022.
4. Qi Sun, Runxin Li, Dingsheng Luo, and Xihong Wu. 2008. Text segmentation with LDA-based Fisher kernel. Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies, pages 269–272.
5. Masao Utiyama and Hitoshi Isahara. 2001. A statistical model for domain-independent text segmentation. In Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, pages 499–506, Toulouse, France.
6. H. Misra, O. Capp'e, and F. Yvon. Using LDA to detect semantically incoherent documents. In Proceedings of CoNLL, Manchester, U.K., 2008

YOUR ESSAY SHOULD HAVE THE FOLLOWING STRUCTURE:

- Introduction: Introduce the subject and state your thesis. The objective is to demonstrate that you understand the essay contest question and have formed a response to it.
- Introduction: Introduce the subject and state your thesis. The objective is to demonstrate that you understand the essay contest question and have formed a response to it.
- Conclusion: Summarize the research and analysis presented in the essay and set forth your recommendations and conclusions. Drawing on ideas already presented, you should demonstrate that your conclusions support the thesis you put forward in the opening paragraphs. Your aim is to convince the reader that your thesis, facts, and analysis are reasonable and valid.

Essays that use a variety of sources—academic journals, news magazines, newspapers, books, government documents, publications from research organizations—fare better in the contest. Web sites and Web pages should not be the only source of information for your essay. When citing Internet sources, include the following information: author(s), title of work, Internet address, and date information was accessed.

YOUR ESSAY MUST:

- address all parts of this year's contest question in English;
- NOT have your name, teacher's name, or school name anywhere on the essay;
- Font must be no smaller than 12 point Times or similarly sized font;
- have a title and page numbers;
- have no more than 2,500 words. The word count limit includes all words in the text but does not include the bibliography and notes (you may exclude the words in the in-text citation from your word count);

№ 3 SIW

Topic: Create a UML model for LDA text segmentation

The purpose of the assignment: To be able to apply the knowledge in practice acquired by the student during the lesson.

Submission form: Writing and oral

Assignments: Based on the result of the SIW2, to design a UML model for this work.

Methodical instructions: Proper use of Internet resources.

References:

1. Brants, T., Chen, F., & Tsochantaridis, I. (2002). Topic-based document segmentation with probabilistic latent semantic analysis. CIKM'02 (pp. 211–218). Virginia: ACM.
2. Doug Beeferman, Adam Berger, and John Lafferty. 1999. Statistical models for text segmentation. Machine learning, 34(1):177–210.
3. David M. Blei, Andrew Y Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. Journal of Machine Learning Research, 3:993–1022.
4. Qi Sun, Runxin Li, Dingsheng Luo, and Xihong Wu. 2008. Text segmentation with LDA-based Fisher kernel. Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies, pages 269–272.

5. Masao Utiyama and Hitoshi Isahara. 2001. A statistical model for domain-independent text segmentation. In Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, pages 499–506, Toulouse, France.
- 6.H. Misra, O. Capp'e, and F. Yvon. Using LDA to detect semantically incoherent documents. In Proceedings of CoNLL, Manchester, U.K., 2008
7. "ISO/IEC 19501:2005 - Information technology - Open Distributed Processing - Unified Modeling Language (UML) Version 1.4.2". Iso.org. 1 April 2005. Retrieved 7 May 2015.

YOUR Design SHOULD HAVE THE FOLLOWING STRUCTURE:

The UML model will include these diagrams:

- Use-case diagram
- Class diagram
- Object diagram
- Communication diagram
- Sequence diagram

You will must to describe the role of every diagram, and it's process of work.

Essays that use a variety of sources—academic journals, news magazines, newspapers, books, government documents, publications from research organizations—fare better in the contest. Web sites and Web pages should not be the only source of information for your essay. When citing Internet sources, include the following information: author(s), title of work, Internet address, and date information was accessed.

YOUR ESSAY MUST:

- address all parts of this year's contest question in English;
- NOT have your name, teacher's name, or school name anywhere on the essay;
- Font must be no smaller than 12 point Times or similarly sized font;
- have a title and page numbers;
- have no more than 1,500 words. The word count limit includes all words in the text but does not include the bibliography and notes (you may exclude the words in the in-text citation from your word count);