**Идеи за курсови проекти: 2021**

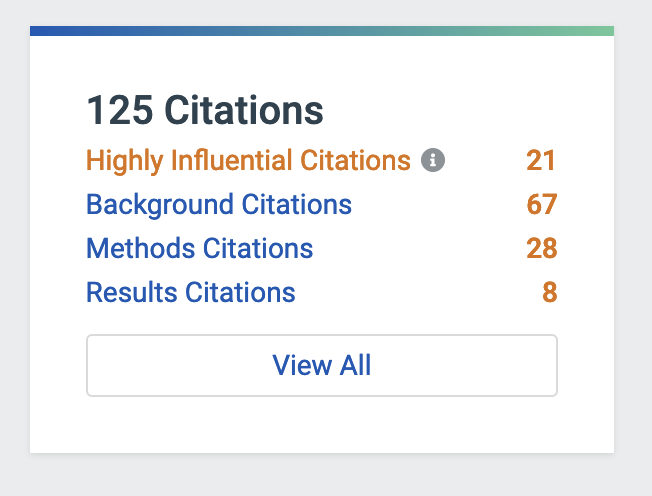
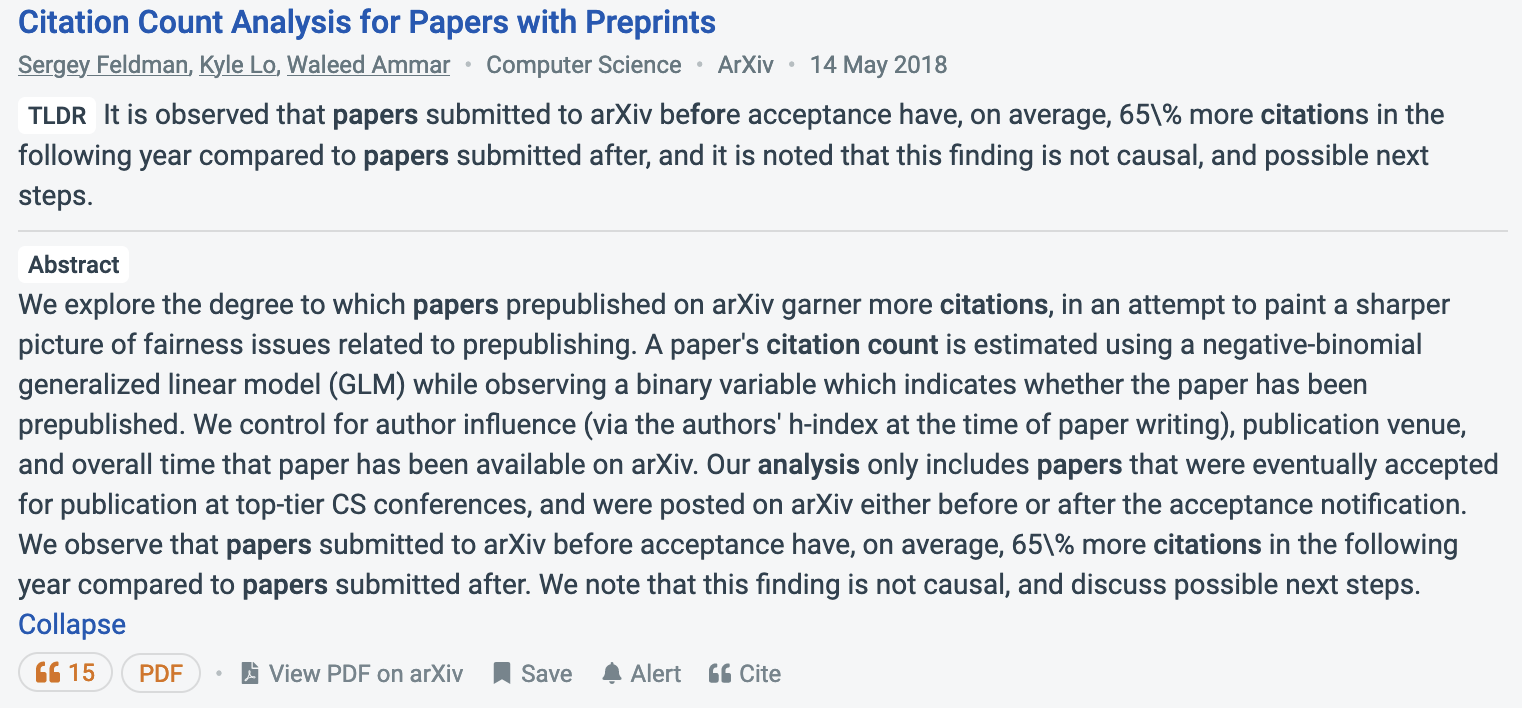
**2021-11-18:**

* Пропаганда в български/македонски мемета
* **Проверяване на твърдения за картинки**
  + Fact-checking claims about images
    - <https://aclanthology.org/D19-1216/>
      * Who cites the above: <https://scholar.google.com/scholar?cites=10191023313524116282&as_sdt=2005&sciodt=0,5&hl=en>
    - <https://www.snopes.com/fact-check/satellite-image-diwali/>
      * Task: “A NASA satellite photograph shows India during Diwali celebrations.” + image → true/false
      * Better formulation:
        + Work with the claim from the original tweet: <https://twitter.com/NASAhistory/status/1456387292864581636>
        + Данни от Иван/Момчил: извличане на двойки от туитове с твърдения за картинки и съответстваща статия от фактчекващ уебсайт
      * Задачи:
        + Проверка чрез сравняване с вече проверявани твърдения
        + Истинска проверка:

<https://gitlab.com/didizlatkova/fake-image-detection/>

Reverse image search: does it appear on bad websites, …

Was the image manipulated, ...

* + - <https://www.snopes.com/fact-check/category/photos/>
    - “Where are the facts”:
      * Paper:<https://www.aclweb.org/anthology/2020.emnlp-main.621/>
      * Previous Paper:<https://arxiv.org/pdf/1910.02202.pdf>
      * Previous Previous Paper:<https://github.com/nguyenvo09/CombatingFakeNews>
      * Data:<https://github.com/nguyenvo09/EMNLP2020>
  + Cross-lingual
    - <https://www.poynter.org/ifcn-covid-19-misinformation/>
    - 7,519 fact checking efforts from 95 fact checking organizations in 136 countries and regions
    - For each fact check, the dataset includes when the fact checker saw the specific claim, the fact checking organization that fact checked the claim, the countries where the fact check was disseminated, and a description of the fact check (all descriptions are in English, regardless of the fact checking organization’s country).
  + System to monitor disinformation in social media
* **Разпознаване на говор: система за факт-чекване в реално време, като се сравнява с проверени факти**
  + checkworthiness
  + previously fact-checked claims
    - <https://www.aclweb.org/anthology/2020.acl-main.332/>
    - CLEF-2021 CheckThat! Lab: <https://gitlab.com/checkthat_lab/clef2021-checkthat-lab/-/tree/master/task2>
    - Facebook XLS-R
    - <https://arxiv.org/abs/2111.09296>
    - <https://ai.facebook.com/blog/xls-r-self-supervised-speech-processing-for-128-languages>
* **Analysing scientific publications**
  + What makes a good publication (impact factor / citations)
  + Abstract/Title generation based on the rest of the paper
  + Identifying key (important) contributions of a publication
  + <https://research.semanticscholar.org/publications>
  + <https://www.semanticscholar.org/paper/Identifying-Meaningful-Citations-Valenzuela-Ha/1c7be3fc28296a97607d426f9168ad4836407e4b>
  + Semantic Scholar offers statistics about the paper:
    - 
  + TLDR from SS
    - ​​
  + Generate layman summary like in Science Daily
    - <https://www.sciencedaily.com/releases/2021/11/211116103110.htm>
    - <https://www.semanticscholar.org/paper/We-Can-Explain-Your-Research-in-Layman's-Terms%3A-at-Dangovski-Shen/ac4e70290f02dc89b9dfcf6c6dfb7f49fbc6068c>
  + Datasets from SemanticScholar:
    - <https://allenai.org/data?tag=Semantic%20Scholar>
    - <https://allenai.org/data/meaningful-citations>
  + Summarize papers based on citations to them
    - <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.59.2666&rep=rep1&type=pdf>
  + Shared tasks at SDproc.org
    - <https://sdproc.org/2022/sharedtasks.html>
  + Life summarization of an ongoing event
    - <https://www.bbc.com/news/live/uk-59331826>
  + Паралелна формулировка -- Ключови моменти от новина:
    - Източници:
      * <https://www.dnevnik.bg/bulgaria/2021/11/15/4279861_deniat_sled_izborite_produljavame_promianata_oshte/>
      * <https://www.dnevnik.bg/bulgaria/2021/11/15/4279861_deniat_sled_izborite_produljavame_promianata_oshte/>

**2021-11-04:**

* вижте този туториал:
  + <https://propaganda.math.unipd.it/cikm21-tutorial/>
  + Видео: <https://zoom.us/rec/play/x3Iegr3hC3yb8uF1KGyYUHdxpG6_qKavjl3pIcG8mk2Z4goaQ5ZS3GHpr6nGYnrKLrD9zaJr3Fd4cTyg.ufGu8tTjFKw9WkUb>
  + Датасети: <https://arxiv.org/abs/2111.03299>
* Пропаганда в български/македонски мемета
  + За английски:
    - Prta (text only):
      * <https://www.tanbih.org/prta>
    - SemEval-2021 Task 6: <https://github.com/di-dimitrov/SEMEVAL-2021-task6-corpus>
    - Статии
      * <https://aclanthology.org/2021.semeval-1.7/>
      * <https://aclanthology.org/2021.acl-long.516/>
* Проверяване на твърдения за картинки)
  + <https://aclanthology.org/D19-1216/>
* Българско ГЛУ
* Разпознаване на говор

2021-10-26:

* Preslav: check-worthiness
  + лъжешли - сайт (на български дебати с видео) - cross-language
* Здравко(**Fauxtography** ми звучи доста интересно и предизвикателно като идея за дипломна работа, тъй като включва и NLP, и CVision. Същевременно доста полезно, имайки предвид голямата вълна от fake news. Бих се радвал да работя по тази тема :) ):
  + NLP + vision
    - propaganda in memes
    - positive/negative propaganda techniques in memes
    - check-worthiness of claims about images
    - previously fact-checked claims about images
    - fact-checking claims about images
    - hero, vilain, victim on the memes
  + chatbot for informal
  + Автоматично генериране на unity-testing
    - CodeQA: A Question Answering Dataset for Source Code Comprehension
      * <https://arxiv.org/abs/2109.08365>
    - CodeBERT: A Pre-Trained Model for Programming and Natural Languages
      * <https://arxiv.org/abs/2002.08155>
  + Момчил: Можем ли да я свържем с Мартин Вечев?
  + NLP + vision за български
  + propaganda/persuasion detection for memes for Bulgarian (cross-lingual)
  + реклама: дали пасва на контекста
  + Out of context images? Are they from this event?

**1. Българско GLUE**

* KLUE: Korean Language Understanding Evaluation <https://arxiv.org/abs/2105.09680>
* Russian GLUE: <https://russiansuperglue.com/>

2. Previously fact-checked claims cross-lingual, for images

* Cross-lingual
  + <https://www.poynter.org/ifcn-covid-19-misinformation/>
  + 7,519 fact checking efforts from 95 fact checking organizations in 136 countries and regions
  + For each fact check, the dataset includes when the fact checker saw the specific claim, the fact checking organization that fact checked the claim, the countries where the fact check was disseminated, and a description of the fact check (all descriptions are in English, regardless of the fact checkixng organization’s country).
* System to monitor disinformation in social media

3. Propaganda/persuasion detection for memes for Bulgarian (cross-lingual)

4. Check-worthiness of claims about images

5. Out of context images? Are they from this event?

6. Fact-checking claims about images

7. Prompt-based methods / GPT-3

<http://pretrain.nlpedia.ai/>

<https://arxiv.org/abs/2107.13586>

<https://arxiv.org/abs/2009.07118>

8. Positive/negative propaganda techniques in memes

8. Positive/negative propaganda techniques in memes

10. Concerned/skeptical memes

10. NLG with language models for dialogue systems (like DialoGPT but actually working)

11. Generation of fact-checks

12. System to monitor fake claims in Twitter in various languages

13. Generating fact-checkable claims from social media text (threads)

14. Generating non-clickbaity titles from news articles

15. Analysing scientific publications

* What makes a good publication (impact factor / citations)
* Abstract/Title generation based on the rest of the paper
* Identifying key (important) contributions of a publication

16. [По-приложна дипломна работа] Има работа за повече от един човек, но ако го преборим ще е много яко. Project: Инструменти за българско NLP:

* Stemmer/Lemmatiser (Това на Преслав добре ли е или може да се подобри?)
  + Bulstem: <https://github.com/mhardalov/bulstem-py>
  + Lemmatizer: <https://github.com/grigoriliev/BGLangTools> (<https://arxiv.org/abs/1506.04228>)
* POS tagger, NER
  + POS/NER: <https://gate.ac.uk/sale/tao/splitch23.html#x28-52600023.20.3>
* Sentence vectors
* Language models
* Entity Linking
* Spellcheck / Grammatical error correction
* ASR / TTS
* Да се види
  + Stanza <https://stanfordnlp.github.io/stanza/>
  + Гръцки framework: <https://github.com/nlpaueb/gr-nlp-toolkit>