Mehdi Aghakishiyev에서 모두에게: 04:05 PM

HW2 grading, how is novelty graded, what are the criteria for it?

윤서 최에서 모두에게: 04:07 PM

방금 페이지 다시 말씀해주실수 있나요? Humor 다음에 말씀하신거를 제대로 못들었어요 ㅜ

20170004\_강동현에서 모두에게: 04:07 PM

HW3 1. about corpora

Junseop Ji에서 모두에게: 04:07 PM

reddit.com/r/humor 말하시는건가요?

20170004\_강동현에서 모두에게: 04:07 PM

I sent my first message as a private message, so I resent the message

윤서 최에서 모두에게: 04:08 PM

humor 카테고리말고 하나더 있지 않았나요 혹시…?감사합니다 :)

20160413\_ 윤소영에서 모두에게: 04:12 PM

Can we use multiple websites, or multiple corpus, for the used corpus?

장채린에서 모두에게: 04:14 PM

csv output: will utf-32 format be okay?

[TA] 20195238\_Huije Lee에서 모두에게: 04:15 PM

[Notice] Anyone who have questions about report of HW2, please send e-mail. (He cannot participate in office hour.) Sorry for Mehdi Aghakishiyev and Soyeong Yoon.

[TA] Eugene Jang에서 모두에게: 04:17 PM

https://github.com/SKTBrain/KoBERT/blob/master/requirements.txthere's an example of a requirements.txt file

Dhammiko Arya에서 모두에게: 04:17 PM

so are we allowed to use Meriam Webster Dictionary to check relationships between Noun - pronunciation? For example, as Prof mentioned, we can check on Meriam Webster1. Wind, -> wind2. wind, -> wi\_nd

윤서 최에서 모두에게: 04:23 PM

I’m little confuse, can we use twitter data…?I knew that I need to use developer account to crawl twitter data and isn’t this private data or etc…?

강성재에서 모두에게: 04:45 PM

질문을 채팅에 올리면 되나요?

[TA] Eugene Jang에서 모두에게: 04:46 PM

네

강성재에서 모두에게: 04:47 PM

음 먼저 혹시 저번 오피스아워에서 샘플로 쓰이는 corpus가 3개 정도를 언급해주셨다고 했던데, 혹시 이에 대해 알려주실 수 있나요?

DongYeun Lee에서 모두에게: 04:50 PM

한국어로 한번 더 말씀해주실수있나요??

[TA] Eugene Jang에서 모두에게: 04:50 PM

Here are some example general sources that we suggested that could have some heteronym examples: www.reddit.com/r/humor www.reddit.com/r/wordplay

Assem Zhunis에서 모두에게: 04:51 PM

The link doesn’t work..

Jaemin Kim에서 모두에게: 04:53 PM

방금 들어왔는데 혹시 어떻게 진행되는지 여쭤볼수 있을까요?

[TA] Eugene Jang에서 모두에게: 04:53 PM

They work for me? You could try googling "humor subreddit" or "wordplay subreddit"질문 적어주시면 순서데로 대답합니다

강성재에서 모두에게: 04:56 PM

혹시 위에 두 링크가 사용할 만한 sample corpus 인가요?

Jaemin Kim에서 모두에게: 04:56 PM

1. output format에 관련한 질문입니다. pronouncing annotation을 output에 적으라고 하셨는데 저희가 배운 CMUdict 포맷이 아닌 IPA 포맷으로 적어도 될까요?

정빈 손에서 모두에게: 04:57 PM

cmudict를 사용하면서 여러 단어들의 발음을 봤을 때 일반적으로 알고있는 heteronym이 아닌 단어들도 여러가지 발음이 있는 경우가 많던데, 혹시 발음이 다르지만 같은 의미인 경우도 있는건가요?

Seongwoo Shin에서 모두에게: 04:57 PM

방금 얘기하셨던 순서 채팅으로 한번만 올려주실 수 있으신가요?

[TA] Eugene Jang에서 모두에게: 04:59 PM

IPA Format is fine as pronounciation annotations

[TA] 20195238\_Huije Lee에서 모두에게: 05:00 PM

@Jaemin Kim 발음 포맷은 둘다 괜찮습니다.

[TA] Eugene Jang에서 모두에게: 05:02 PM

Yes the two links are sample sources on the internet that can construct your corpusWords with multiple pronunciations for the same meaning exist, but are not the words that we are looking for

JH Lee에서 모두에게: 05:08 PM

단어가 여러번 들어간 문장이 점수를 높이 받아야 한다고 되어 있는데 wind(n)이 10번 들어간 문장과 wind(n) wind(v) tear(n) tear(v) 이 각 한번씩 들어간 문장이 있으면 전자가 더 높은 점수라는 뜻인가요?

20180526 Hyejin Lee에서 모두에게: 05:12 PM

wind로 발음되는 wind(n)과 wind로 발음되는 wind(v)가 같이 있을 때는 homograph 쌍으로는 치지 않는게 맞나요?

강성재에서 모두에게: 05:18 PM

앗 죄송한데, 샘플 corpus를 다시 적어주실 수 있나요?? 채팅이 초기화돼서 확인을 못하네요, 레딧에 humor하고 무슨 word였던 것 같은데 다시 적어주시면 감사하겠습니다

현구 이에서 모두에게: 05:18 PM

외부 dictionary 사용해도 된다고 저번시간에 수업 끝나실때 말씀하신것 같은데 정말 사용 가능한지 확인할 수 있을까요?

[TA] Eugene Jang에서 모두에게: 05:19 PM

reddit.com/r/humor reddit.com/r/wordplay외부 dictionary에서 정보 가져오는거 사용 가능합니다

20170490 Jaejun Lee에서 모두에게: 05:26 PM

혹시 방금 말씀해주신 우선순위 예 텍스트로 올려주실 수 있나요?

[TA] 20195238\_Huije Lee에서 모두에게: 05:27 PM

Geonho Lee 학생 질문에서 첨언하자면, 4가지 파일(docx, pdf, csv, py)에 더해서 추가 파일 제출(requirements.txt, readme and so on)을 해주셔도 됩니다.For comments about Geonho Lee's question, you can submit additional files(requirements.txt, readme and so on).

Jooyon Kim에서 모두에게: 05:31 PM

제가 지금 접속해서 앞에서 이미 설명하셨는지는 잘 모르겠어서 홍시 중복된다면 죄송합니다.Heteronym list를 외부에서 들고 오는게 아니라, nltk cmudict 등을활용해서 제가 직접 코드 안에서 listing을 한 후에 문장들을 분석하기 시작하는건 상관이 없나요?

조석민에서 모두에게: 05:31 PM

외부 라이브러리 사용에 대한 질문입니다 word sense disambiguation을 위해 딥러닝 모델을 사용하는 건 괜찮나요?

JH Lee에서 모두에게: 05:32 PM

단어가 여러번 들어간 문장이 점수를 높이 받아야 한다고 되어 있는데 wind(n)이 10번 들어간 문장과 wind(n) wind(v) tear(n) tear(v) 이 각 한번씩 들어간 문장이 있으면 전자가 더 높은 점수라는 뜻인가요?

Yohan Kim에서 모두에게: 05:35 PM

1. Wind[Wind] / Wind[Waind] / Tear [Tier] / Tear [Tar]2. Wind[Wind] / Wind[Wind] / Tear[Tier] / Tear[Tar]3. Wind[Wind] / Wind[Wind] / Tear[Tier] / Tear[Tier]4. Wind[Wind] / Wind[Waind] / Tear[Tier]5. Wind[Wind] / Wind[Wind] / Tear[Tier]6. Wind[Wind] / Wind[Waind]7. Wind[Wind] / Tear[Tier]제가 필기한 내용 옮긴 내용입니다 (정확한가요?)

Paco Kwon에서 모두에게: 05:35 PM

1. 윈드 와인드 티어 테얼2. 윈드 윈드 티어 테얼3. 윈드 윈드 티어 티어4. 윈드 와인드 티어5. 윈드 윈드 티어6. 윈드 와인드7. 윈드 티어If you can read korean

20160871 chitipat phetmunee에서 모두에게: 05:35 PM

thank you so much

장채린에서 모두에게: 05:36 PM

waɪnd', 'wɪnd'tɛər, tɪərto TA if you need this..

Jooyon Kim에서 모두에게: 05:37 PM

제 질문은 방금 답변으로 해결되었습니다.

Mehdi Aghakishiyev에서 모두에게: 05:38 PM

yes

김재민에서 모두에게: 05:38 PM

저도 모두 해결되었습니다

20170490 Jaejun Lee에서 모두에게: 05:42 PM

우선순위 적어주셔서 감사합니다.rule 3을 보면 같은 품사의 heteronym을 가지는 경우가 다른 품사의 heteronym을 가지는 경우보다 우선순위가 높다고 되어있는데요.wind waind tear > wind wind tear가 맞나요..?

강건희에서 모두에게: 05:43 PM

wind wind 로 쓰이는 경우가 같은 품사인거 아닌가용?

20170490 Jaejun Lee에서 모두에게: 05:44 PM

네 그런데 오늘 말씀하신 우선순위는 rule 3에 반하는 것 같아서요

20170004\_강동현에서 모두에게: 05:45 PM

아까 질문 덜 된거가 있는데요 혹시 Pronouncing Dictionary 사용 가능하나요?

[TA] Junseop Ji에서 모두에게: 05:51 PM

allen입니당

[TA] 20195238\_Huije Lee에서 모두에게: 05:52 PM

앞서 말한 위키토리 예시와 같이 직접적인 heteronym 표기를 사용하지 않는 pronouncing dictionary는 사용 가능합니다.

강건희에서 모두에게: 05:54 PM

manual하게 만든 heteronym list나 오류를 잡기위해 만든 exception list 를 써도 되나요?

20170278\_Jongchan Park에서 모두에게: 05:55 PM

오피스 아워중에 heteronym 인 단어가 많은 문장이 우선순위라고 하셧는데 그러면 wind1 wind2 (서로 다른발음)이 들어간 문장이 bank라는 homograph가 10개 들어간 문장보다 우선순위 라는건가요?

김주연에서 모두에게: 05:56 PM

제가 딕셔너리 openAPI를 사용했는데, 하루에 할수있는 리퀘스트 수가 정해져있는데 채점하실때 혹시 저희 코드 여러번 돌리시나요? 딱한번까지만 제 코드가 잘 돌아갈거여서요.

김재민에서 모두에게: 06:02 PM

저 마지막 질문 해도될까요??제가 지금 알바중이라 채팅으로 가능할것같습니다죄송합니다..

[TA] Eugene Jang에서 모두에게: 06:03 PM

질문 해주세요

김재민에서 모두에게: 06:03 PM

wiktionary에서 계속해서 parse하는 방식으로 할때 조교님께서 크롤링 코드를 쓰고 크롤링한 정보를 따로 저장해서 사용하라고 하셨는데, 그렇게 하명 너무 오랜시간이 걸릴것같아서 필요한 단어를 그때그때 찾는 크롤링 함수를 구현하고 이를 사용해 찾아 저장한다음에 나중에 중복 단어가 등장할때는 그대로 사용하는 방법을 채택해도되나요?