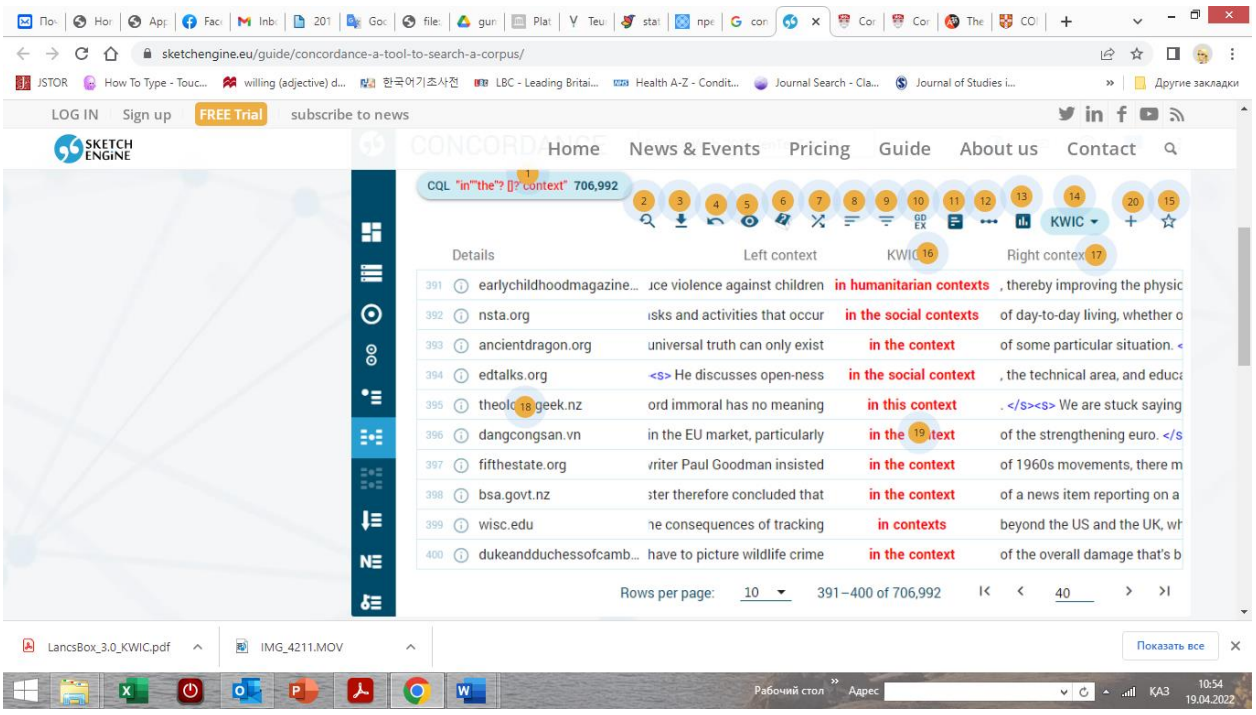
Дәріс №12
Дәріс тақырыбы: Корпус менеджерлері
Сұрақтар:

1. Шығарылым интерфейстері
2. Лингвистикалық емес корпустардың корпус менеджерлері

1. Корпус менеджерлеріндегі іздеу нәтижелері әдетте конкорданс түрінде ұсынылады. Жазушының кітабында немесе шығармасында қолданылатын әліпби тәртібіндегі сөздердің қай жерде және қандай сөйлемдерде кездесетіні туралы мәліметтері бар кітап немесе құжат:

Конкорданс - бұл көптеген іздеу нұсқалары бар ең қуатты құрал. Ол сөздерді, сөз тіркестерін, статистиканы, құжаттарды, мәтін түрлерін немесе корпус құрылымын таба алады және нәтижелерді контексте конкорданс ретінде көрсете алады. қажетті нәтижеге қол жеткізу үшін Конкордансты сұрыптауға, сүзуге, санауға және одан әрі өңдеуге болады. Бұл ең қуатты құрал болса да, ірі корпустар қолданатын конкорданс көптеген нәтижелер бере алады, сондықтан оларды талдау мен түсіндіру қиын болуы мүмкін. Көру параметрлері лемма, тегтер және басқа атрибуттар, мәтін түрі (метадеректер) немесе корпус құрылымы сияқты қосымша ақпаратты көрсетуге мүмкіндік береді. «Қосымша» вкладкасындағы SQL іздеу белгілі бір критерийлер немесе таңдау критерийлер бойынша күрделі іздеу үшін қолданылады. https://www.sketchengine.eu/guide/concordance-a-tool-to-search-a-corpus/

Қараңыз:



2. Лингвистикалық емес корпустардың корпус менеджері дегенде интернет желісі түсініледі. Вед кеңістіктің өзі үлкен көптілді корпус ретінде қаарстырыла алады дейді Захаров пен Богданова. Себебі интернетте өте үлкен көлемдегі машинада өңдеуге болатын деректер қор ыбар. Корпустар жасаудың деректері ретінде веб кеңістіктегі түрлі мәтіндер алынады.. Веб-кеңістікті корпус ретінде пайдалану кезінде іздеу жүйелері корпус менеджерлерінің рөлін орындайды. Интернетте кітапхана каталогтарын еске түсіретін классификациялық типтегі жүйелер бар (каталогтар, орысша жалпы атауы – «каталог-анықтамалары»/«каталоги-справочники»). Интернетте ақпаратты іздеудің негізгі құралы бүкіл интернет кеңістігін индекстейтін вербалды типтегі ғаламдық ақпараттық іздеу жүйелері (іздеу машиналары - search engines) болып табылады. Вербалды түрдегі негізгі

іздеу жүйелеріне (ең алдымен деректер қорының көлемі бойынша) мыналар жатады: Google, Fast Search (AllTheWeb), AltaVista, WiseNut, HotBot, MSN Search, Teoma. Ресейлік жүйелердің ішінде негізгілері үшеу: Яндекс (Яндекс, Яндекс), Рамблер (Rambler), Апорт! (Апорт).

Кез келген іздеу жүйесінің құрамында үш негізгі бөлік бар:

1. Робот
2. Поисковая база данных – так называемый индекс
3. Поисковая система (В.Захаров, С.Богданова)

Ақпараттық сұрау - бұл белгілі бір ақпарат қажеттілігінің ауызша көрінісі. Сұраулар пән және формальды мазмұны бойынша талданады және корпуспен жұмыс істейтін қолданбалы бағдарламаның сұраныс тілінің сөздік құрамы тұрғысынан сипатталады. Сұрау салуды сіздер сабаққа дайыналу барысында немесе басқа да мақсаттарда жиі пайдаланасыздар. Мысалы гуглдің сұрау салу жүйесі. Жалпы, сұрау тілі үлгісі келесі элементтерді қамтиды: 7 (оқыңыз)

Кітаптарыңда **Bonito/Manatee** сауал тілі жан жақты берілген. Manatee - корпус менеджері. Bonito - бұл Manatee корпус менеджерінің графикалық пайдаланушылық интерфейсі (GUI). Бұл жүйені Чехияда Масарика атындағы университеттің информатика факультетіндегі NLPlab (Natural Language Processing Laboratory) лабораториясы мен П. Рыхли жасаған.

Мен сіздерге Скетч енжинді таныстырайын.

Sketch Engine is a corpus manager and analysis software has developed by Lexical Computing since 2003. This software consists of three main components which enable to search and build text corpora.

- **Bonito** – a graphical user interface to corpora maintained, see the changelog of Bonito
- **Manatee** – a corpus management tool including corpus building and indexing, fast querying and providing basic statistical measures
- **FinLib** – fast indexing library, see the changelog of FinLib

https://www.sketchengine.eu/documentation/manatee-changelog/

Bonito жүйесінің негізгі ерекшеліктері (кітаптан); Сауалдар (кітапта), сауал типтері (кітапта), шаблондар (кітапта); сауал мысалдары 7 (кітапта)

Мынадай белгілі әмбебап корпус менеджерлері бар: SARA, XAIRA (BNC), Manatee/Bonito, CQP, DDC. Корпус деректерін өңдеу үшін дерекқорды басқару жүйелері (систем управления базами данных - СУБД) немесе іздеу жүйелері негізінде менеджерлер әзірленуі мүмкін. Оған мысалы орыс тілінің ұлттық корпусының іздеу жүйесі. Мұнда іздеу Яndex.Server 3.8 Professional іздеу жүйесі арқылы іске асады.