|  |  |
| --- | --- |
|  |  |

**ДИПЛОМСКА РАБОТА**

**Тема**

**Користење на генеративни модели за автоматско создавање на SQL прашања користејки природен јазик**

|  |  |
| --- | --- |
| Ментор | Изработил |
| проф. др.  Димитар Трајанов | Павел Стојановски Андоновиќ, индекс бр. 175022  e-mail:pavel.gjorgovski@students.finki.ukim.mk |

*Скопје, Октомври 2023*

**Содржина**

[Апстракт 1](#_Toc223681204)

[1 Вовед 2](#_Toc223681205)

[2 Структура на семинарската 2](#_Toc223681206)

[3 Тело на семинарската 2](#_Toc223681207)

[3.1 Уредување на семинарската 2](#_Toc223681208)

[3.2 Наслови 2](#_Toc223681209)

[3.3 Главниот текст 3](#_Toc223681210)

[3.4 Слики и табели 3](#_Toc223681211)

[3.5 Фусноти (забелешки) 4](#_Toc223681212)

[4 Референци 4](#_Toc223681213)

[5 Проверка на граматичката исправност 5](#_Toc223681214)

[6 Големина и обем на семинарската работа 6](#_Toc223681215)

[7 Предавање на семинарската работа 6](#_Toc223681216)

[8 Презентација 6](#_Toc223681217)

[9 Оценување 6](#_Toc223681218)

[10 Листа за проверка 6](#_Toc223681219)

# Апстракт

*Технологијата за големи јазични модели напредува со неверојатна брзина. Дел од обврски на програмерите можат да бидат поедноставени и автоматизирани со нејзиното користење. Еден од предизвиците со кои секојдневно се соочуваат податочните инженери е создавањето на добро SQL прашање. Во оваа дипломска работа целта е креирање на алатка која ќе овозможи поставување на барање на податочно ниво со користење на природен јазик, кое барање потоа автоматски ќе биде преведено во SQL прашање. Главната примена на ова решение е нетехнички лица да можат да поставуваат SQL прашања без потреба од техничкото познавање на јазикот, како и олеснување на секојдневната работа на податочните инженери.*

# Вовед

Вештачката интелигенција има направено големи напредоци во минатите години, особено во обработка на природни јазици. Генеративни модели, како ChatGPT, се имаат истакнато како моќни алати во голем број домени [1][2][3], дополнително подржани од различни методи на насочување на генеративните модели за да се добие резултат каков што корисникот очекува. Еден од секторите во кои има голем напредок е автоматско генерирање на код [4], и се очекува дека голем дел од шаблонското пишување код ќе биде многу олеснето користејќи алатки од генеративни модели.

SQL е програмски јазик

Во овој труд ќе се разгледа примената на ChatGPT во генерирање на SQL прашања

# Референси

1. Menekşe, M. (2023). Envisioning the future of learning and teaching engineering in the artificial intelligence era: opportunities and challenges. Journal of Engineering Education, 112(3), 578-582. <https://doi.org/10.1002/jee.20539>
2. Malik Sallam, Nesreen A. Salim, Muna Barakat and Ala’a B. Al-Tammemi (2023) - ChatGPT applications in medical, dental, pharmacy, and public health education: A descriptive studyhighlighting the advantagesand limitations
3. Michael Dowling, Brian Lucey, ChatGPT for (Finance) research: The Bananarama Conjecture, Finance Research Letters, Volume 53, 2023
4. Haoye Tian, Weiqi Lu, Tsz On Li, Xunzhu Tang, Shing-Chi Cheung, Jacques Klein, Tegawendé F. Bissyandé (2023) - Is ChatGPT the Ultimate Programming Assistant -- How far is it?

Откако ќе се внесе референцата на крајот на текстот треба да се направи соодветно референцирање во склоп на текстот кој го пишувате. Секаде каде што се користат факти, текст, или информации од некој надворешен извор мора да се стави референца, и на тој начин да се нарави јасна разлика што е ваше а што сте земале како информација од надворешен извор.

На пример ако ставате информации за потрошувачката на енергија кај различни уреди тога треба да го направите на следниов начин:

*“Така на пример денешните notebook компјутери просечно трошат од 5-10 W додека активниот безжичниот мрежен интерфејс просечно троши околу 1.5 W, што прави од 15% до 30% од вкупната потрошувачка на енергија [4].”*

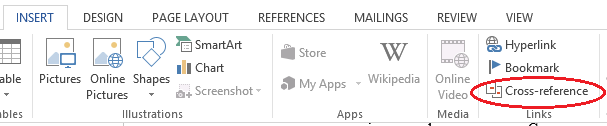
Овој факт е прочитан од некој текст, па за да може да се потврди неговата валидност задолжително е потребно да се стави референца.

Кај деловите од текстот кој претставуваат превод од некоја книга, интернет страница или некој научен труд, референцата треба да се стави само еднаш после првиот параграф.

## Вметнување на референците во Word

Текст едиторите како Microsoft Word овозможуваат едноставно управување и ставање на референци во текстот. Откако ќе ги ставите референците на крајот на текстот како листа со броеви, во текстот референцата се става на следниов начин:

1. Се избира Reference->Cross-Reference . или се оди во панелот INSERT и се избира Cross-Reference
2. Се покажува дијалог како на Слика 3
3. Се избира Reference type: Numbered item и Insert reference to: Paragraph number
4. Некаде на крајот од листата се референците
5. Се избора соодветната референца и се вметнува иста во текстот на соодветното место.



Слика 4 Избор на Cross-Reference кај Word 2010/2013

Graphical user interface, text, application, email

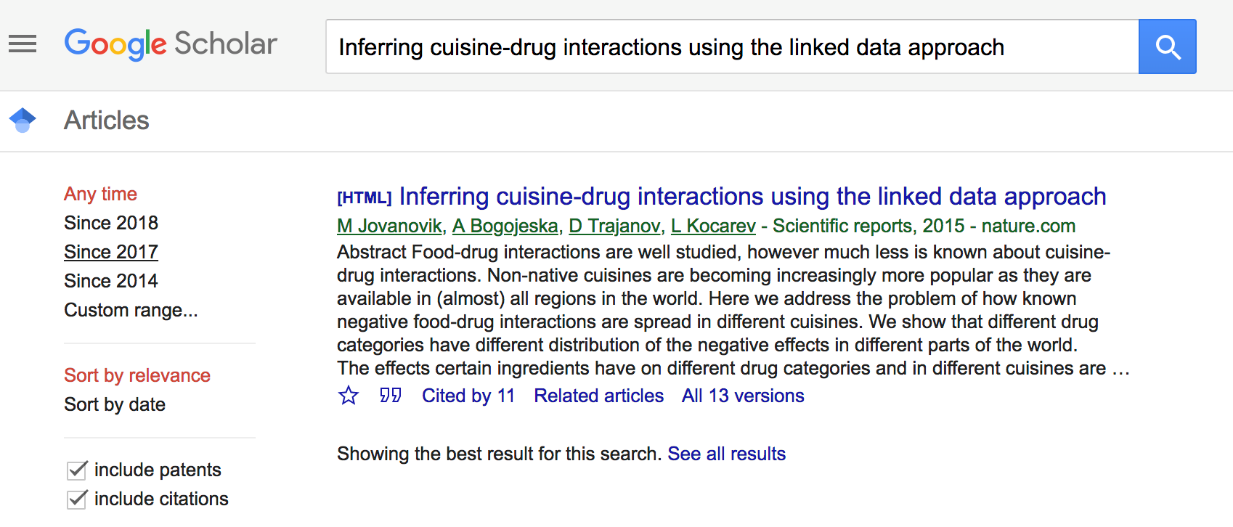
Description automatically generated

Слика 5 Начин на вметнување на референца

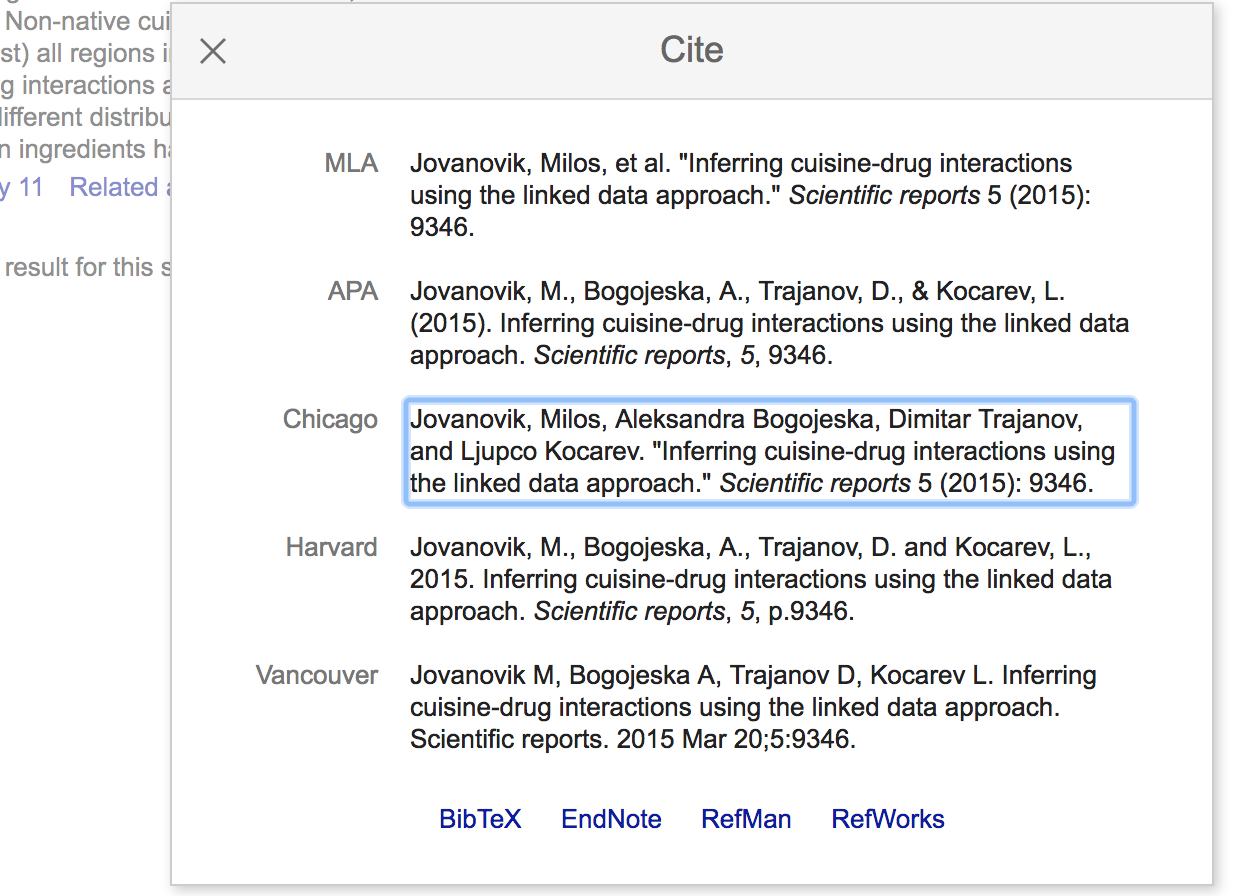
## Средување на референците во соодветен формат

Средувањето на референците во соодветен формат најлесно може да се направи преку користење на google Scholar сервисот. Постапката би била следна

1. Отидете на страницата scholar.google.com
2. Во полето за пребарување внесете го насловот на трудот за кој сакате да го добиете форматот за референца. На пример: “Inferring cuisine-drug interactions using the linked data approach”
3. Пронајдете го соодветниот резултат, кликнете на иконата со слика од наводници **”**



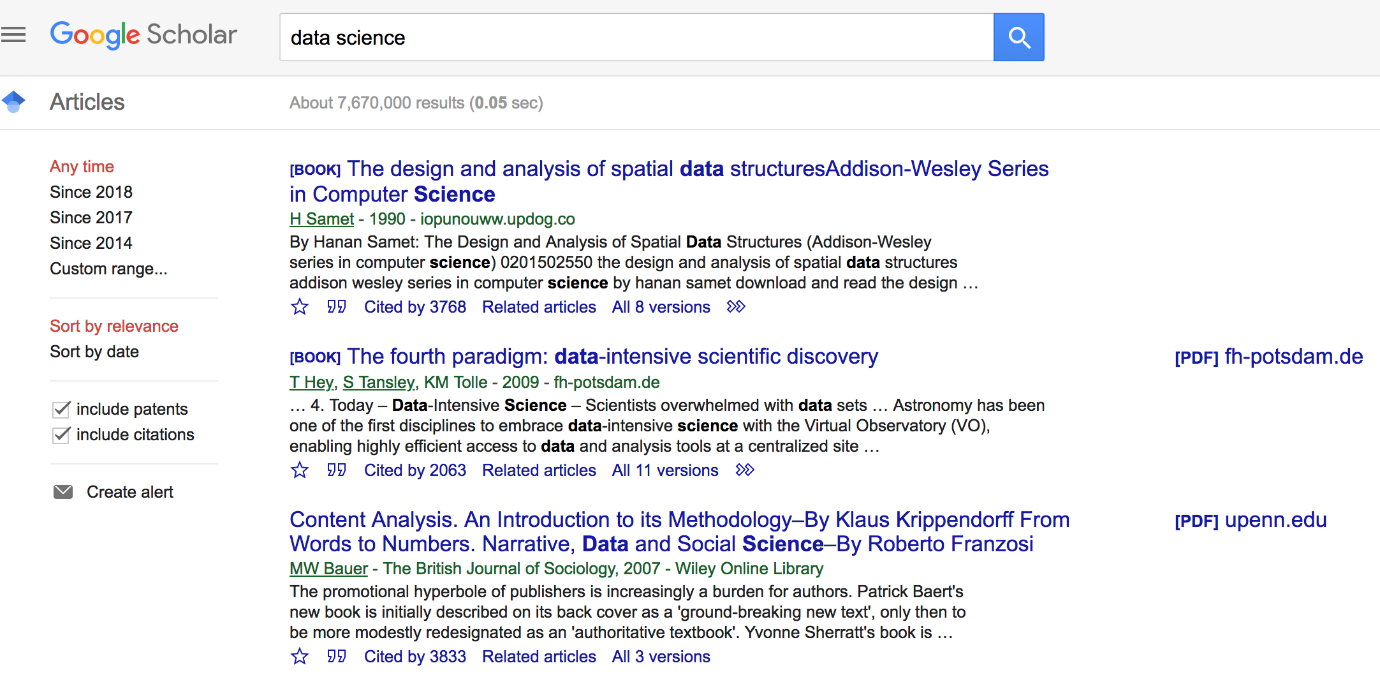
1. По ова ќе ви се појави следниов екран



1. Селектирајте ја референцата од форматот Chicago, копирајте ја и ставете ја во делот за референци во вашиот труд.

## Наоѓање на трудови

Еден од главните портали каде можете да пребарувате научни трудови е scholar.google.com. На оваа станица е пишуваат соодветните клучни зборови и се добива листа на научни трудови кои се поврзани со соодветните клучни зборови. Резултатите кои се добиваат се продадени по значење на трудовите, па така најпрво ќе ги добиете трудовите кои имаат најмногу цитираност.



За да ги добиете само најновите трудови од десната страна на резултатите имате можност да ги филтрирате резултатите. Па така ако одберете Since 2018 ќе ги добиете само трудовите кои се публикувани во последните пет години, или ако одберете Since 2021 ќе ги добиете само трудовите од последната година.

Стратегијата за пребарување би била редоследно да се видат прво сите трудови, пота тие од последните пет години и на крај трудовите од последната година.

Голем број на трудови во последните години со open access лиценца која овозможува бесплатно симнување. Ако за трудот нема линк во Scholar, тогаш можете да се обидете да пребарате директно на Google. Многу често авторите сами ги објавуваат трудовите на нивните веб страници па може да се симнат од таму.

# Проверка на граматичката исправност

Пред да ја предадете вашата семинарска/дипломска работа прво проверете ја од граматички грешки. Во новите верзии на Word со самата инсталација можете да одберете да се инсталира и граматичка проверка за Македонски јазик. Кај постарите верзии треба да инсталирате соодветен пакет за Office 2003/2007/2010 со јазична меѓу врска на македонски јазик. Овој пакет може бесплатно да се симне од овој линк (<http://www.microsoft.com/downloads/details.aspx?FamilyID=ccf199bc-c987-48f5-9707-dc6c7d0e35d0&displaylang=mk>)

Откако ќе го инсталирате, spelling опцијата ќе работи и за Македонскиот јазик.

# Користење на англиски зборови

Треба да се направи напор сите зборови да се преведат на македонски јазик. Во текстовите не смее да се користат англиски зборови напишани на кирилица. Ако е потребно да се користи некој англиски збор тогаш тој се пишува на латиница во оригинал. На пример не се пишува фрејмворк туку рамка или framework.

# Големина и обем на дипломската/семинарската работа

Големината зависи од предметот но генерално е:

* Минимум 30 страници за дипломска работа
* Околу 30 страници по студент (9000 – 15000 зборови на 30 страници) за постдипломските (Магистерски) студии
* Околу 15 страници по студент (4500 – 75000 зборови на 15 страници) за додипломските студии

# Предавање на дипломската/семинарската работа

Дипломската работа се предава со согласност со упатствата дадени на <http://diplomski.finki.ukim.mk/>

Семинарската работа се предава во електронски облик.

# Презентација

Презентацијата треба да биде направена во Power Point, Google Slides или слична апликација. Бројот на слајдови е:

* 10-20 за дипломска со 15-20 минути за презентирање
* 30-40 за постдипломските студии, 15-20 минути за презентирање
* 15-20 за додипломските студии, 5-10 минути за презентирање

# Оценување

За семинарската работа се оценуваат следниве работи:

* Структура на семинарската работа
* Исправност
* Комплетност
* Содржина
* Референци
* Квалитет
* Доставување во рок (за оние семинарски кои имаат временско ограничување)
* Оригиналност
  + Секоја семинарска работа ќе биде тестирана на Google
  + Цитирање и референцирање е прифатливо
  + Комплетно препишување не е прифатливо

# Листа за проверка

Пред да ја предадете дипломската/семинарската работа **задолжително** проверете ги следниве работи:

* Имате насловна страница на која се напишани вашето име и презими и бројот на индекс
* Има содржина
* Има апстракт
* Има вовед
* Има заклучок
* За означување на насловите се користени стилови од групата Heading
* Текстот е напишан со македонска поддршка
* Проверена е граматичката исправност на текстот со алатката за spell-checker за Македонскиот јазик
* Секоја слика има наслов кој се наоѓа под сликата
* Секоја табела има наслов кој се наоѓа пред табелата
* На крај имате листа со референци кои се користеле при изработката на трудот
* Цитирањето и прераскажувањата од користените референци се соодветно обележани до [xx] синтаксата
* Имате презентација за трудот во PowerPoint