

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМ. ІГОРЯ СІКОРСЬКОГО»
ТЕПЛОЕНЕРГЕТИЧНИЙ ФАКУЛЬТЕТ

ФОРМАЛІЗАЦІЯ ПРЕДСТАВЛЕННЯ ЗНАНЬ ЗА СЕМАНТИКОЮ НЕСТРУКТУРОВАНОГО ТЕКСТУ

Магістрант групи ТІ-61м
Науковий керівник

Савченко Микита Миколайович
Крячок Олександр Степанович

МЕТА ДОСЛІДЖЕННЯ

- Створення семантичної моделі представлення знань, вилучених із текстів, яка задовольняла б наступним критеріям:
 - ефективність роботи на великих масивах даних;
 - автономність роботи;
 - вилучення лише найбільш релевантних знань по відношенню до контексту пошуку;
 - гнучкість та можливість розширення представлення знань.

ОБ'ЄКТ ДОСЛІДЖЕННЯ

- Визначення закономірностей за інформацією з неструктурованого тексту

ПРЕДМЕТ ДОСЛІДЖЕННЯ

- Моделі представлення знань для обробки машинописних текстів природної мови та алгоритми формалізації представлення знань і побудови мереж знань

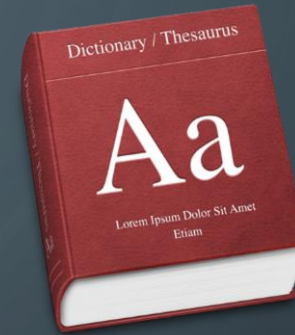
НАЙБІЛЬШ СУТТЄВІ НАУКОВІ РЕЗУЛЬТАТИ

МАГІСТЕРСЬКОЇ ДИСЕРТАЦІЇ

- Розроблена комплексна модель представлення знань на основі неоднорідних семантичних мереж і алгоритм для побудови семантичної мережі знань у автоматичному режимі, який дозволяє будувати мережі знань на основі лише найбільш змістовної інформації, вилученої із текстів
- Набув подальшого розвитку підхід до вилучення знань з використанням алгоритмів Part of Speech Tagging, TF-IDF та неоднорідних семантичних мереж

МЕТОДИ КОНЦЕПТУАЛІЗАЦІЇ ТА ФОРМАЛІЗАЦІЇ ЗНАНЬ

- Методи, засновані на статистичних оцінках
- Методи, засновані на словниках або наперед визначених особливостях природної мови
- Методи, що використовують стійкі конструкції мови для визначення концептів та відношень між ними в текстах
- Методи, що використовують штучний інтелект, алгоритми Machine Learning та нейронні мережі



ПРОБЛЕМИ ІСНУЮЧИХ МЕТОДІВ

- Вузька спеціалізація рішень та алгоритмів
- Необхідність втручання експертів у систему
- Складність вилучення корисної (релевантної) інформації при роботі на великих даних
- Закритість деяких рішень і методів покладених в їх основу (комерційні рішення)



МОДЕЛІ ПРЕДСТАВЛЕННЯ ЗНАНЬ

- Логічні
 - Предикативна логіка
 - Формальні представлення знань
- Мережеві
 - Семантичні мережі
 - Фрейми
- Нечітка логіка
- ...



INTERSYSTEMS IKNOW

Ink

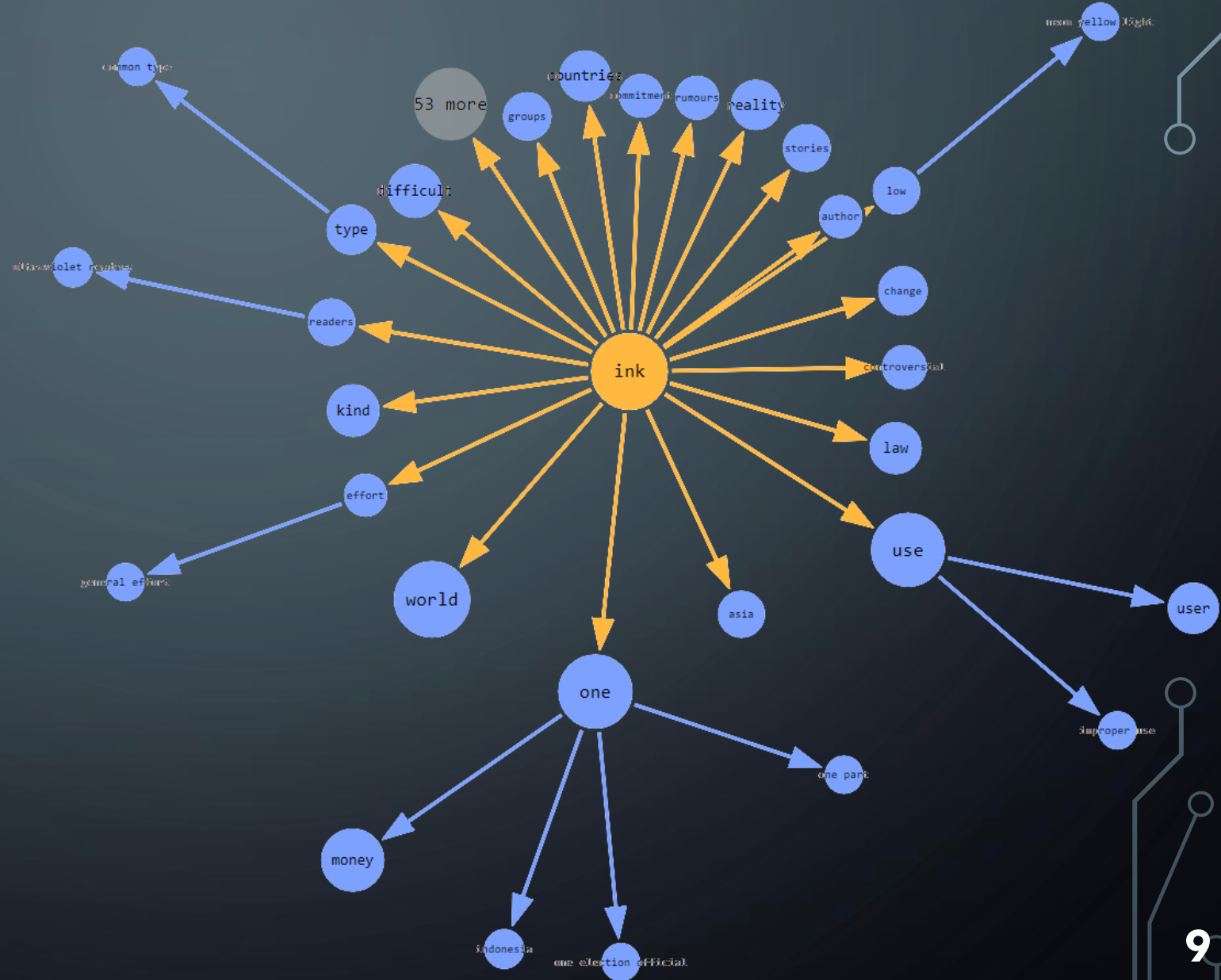
World

One - money

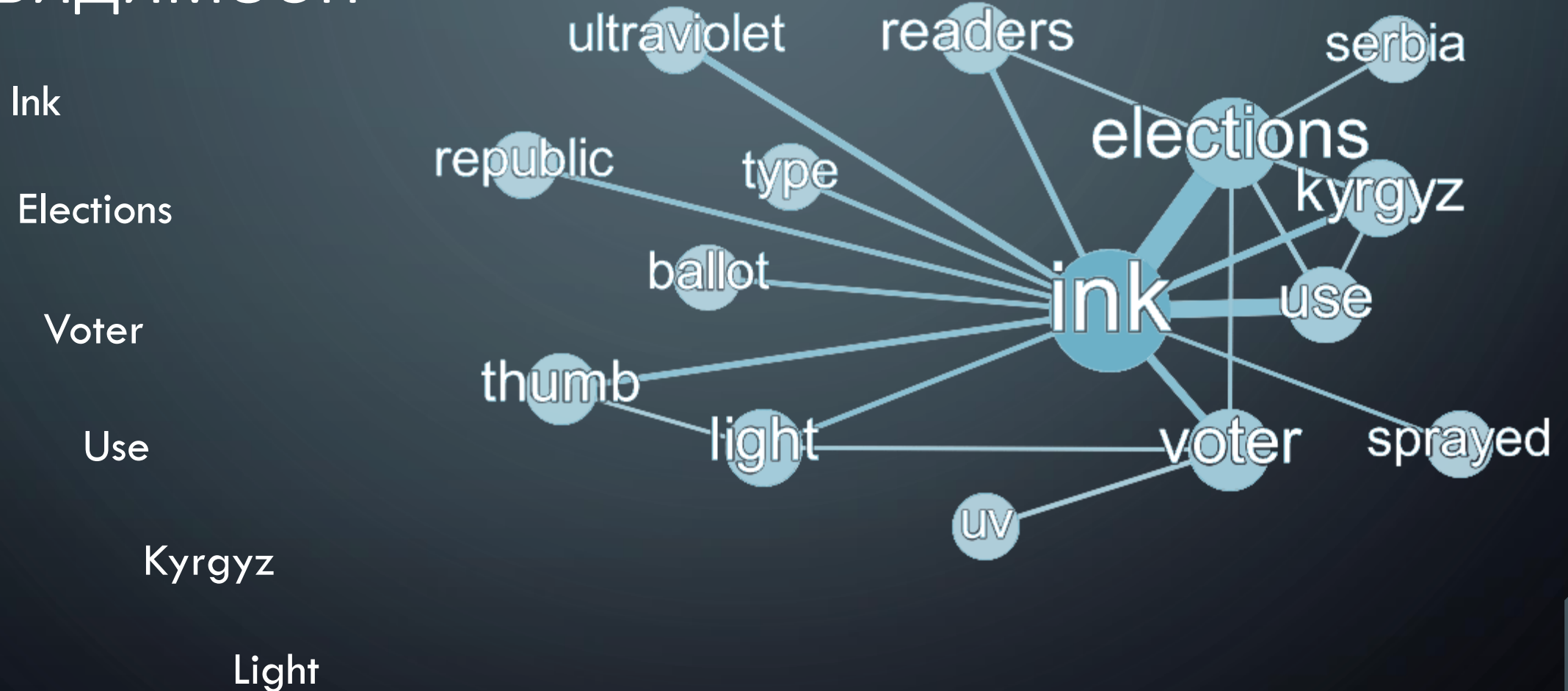
Improper use

Difficult

Countries



ЗАСТОСУВАННЯ ГРАФІВ ГОРИЗОНТАЛЬНОЇ ВИДИМОСТІ



Ink

Elections

Voter

Use

Kyrgyz

Light

РОЗРОБЛЕНИЙ КОМПЛЕКСНИЙ АЛГОРИТМ

- Ідентифікація семантики на основі нейронних алгоритмів Part-of-Speech Tagging
 - Використана нейронна мережа LSTM, на вході до якої подається текст, в якому виконується класифікація слів за частинами мови
- Автоматичне визначення відношень між концептами з довільною типізацією
 - Відношення між концептами в семантичній мережі представлені у такому ж вигляді, що і в тексті без спотворення
- Оцінка «важливості» концептів по відношенню до їх контексту
 - Виконується оцінка концептів з використанням метрик TF-IDF
- Побудова семантичної мережі на основі довільних правил високорівневої граматики
 - Розроблена граMATика для англійської мови, що заснована на роботі з класифікованими частинами мови



КОМПЛЕКСНИЙ АЛГОРИТМ

Ink is harmful radioactive

Ink helps drive democracy

Ink has been controversial

Former soviet republic
is using invisible ink

Regime changing
presidential election

Afghanistan showed
improper use



ВІДБІР КОНЦЕПТІВ, ЩО ПОТРАПЛЯЮТЬ ДО ГРАФУ

Turkey turns on the economic charm three years after a grueling economic crisis. Turkey has dressed its economy to impress. As part of a charm offensive - ahead of 17 December, when the European Union will decide whether to start entry talks - Turkey's economic leaders have been banging the drum to draw attention to recent achievements. The economy is growing fast, they insist. Education levels among its young and large population are rising. Unemployment levels, in percentage terms, are heading fast towards single digits. Inflation is under control. A new law to govern its turbulent banking system is on the cards. The tourism industry is booming and revenues from visitors should more than double to \$ 21bn (£10.8bn) in three years. Moreover, government spending is set to be frozen and a burdensome social security deficit is being tackled. Income and corporate taxes will be cut next year in order to attract \$ 15bn of foreign investment over the next three years. A loan restructuring deal with the International Monetary Fund (IMF) is pretty much in the can. And following recent macroeconomic restructuring efforts, its currency is floating freely and its central bank is independent. The point of all this has been to convince Europe's decision makers that rather than being a phenomenally costly exercise for the EU, allowing Turkey in would in fact bring masses of economic benefits. "The cake will be bigger for everybody," said Deputy Prime Minister Abdullahi Sener earlier this month. "Turkey will not be a burden for the EU budget." If admitted into the EU, Turkey would contribute almost 60n euros (\$ 8bn; £8n) to its budget by 2014, according to a recent impact study by the country's State Planning Organisation. As Turkey's gross domestic output (GDP) is set to grow by 6% per year on average, its contribution would rise from less than 5bn euros in 2014 to almost 20n euros by 2020. Turkey could also help alleviate a labour shortage in Old Europe once its population comes of age. By 2014, one in four Turks - of about 18 million people - will be aged 14 or less. "A literate and qualified Turkish population," insisted Mr Sener, "will make a positive impact on the EU." This runs contrary to the popular view that Turkey is getting ready to dip deep into EU taxpayers' wallets. However, Turkey's assertions are confirmed by Brussels' own impact studies, which indeed say that Turkish membership would be good news for the EU economy. But only over time. Costs are projected to be vast during the early years of Turkey's membership, with subsidies alone estimated to exceed 16.7bn euros and, according to some predictions, balloon to 33.7bn euros. This would include vast agricultural subsidies and regional aid; though such payments should decline as the country's farm sector, which currently employs one in three Turks, would employ just one in five by 2020. Such high initial expenses would be coupled with risks that the benefits flattered up by Turkey's government would never be delivered, say those who feel the Turkish project should be shunned. Some fear that rather than providing an educated, sophisticated labour force for Europe at large, the people who will leave Turkey to seek work abroad will be poor, uneducated and plentiful. More recently, less palatable concerns - at least in liberal European circles - have been voiced, with senior EU or member state officials talking darkly of a "river of Islam", an "oriental" culture and a threat to Europe's cultural richness. Of course, many opponents are politically motivated - their views ranging from xenophobic prejudices about the country's Muslim traditions to well-documented concerns about the government's human rights record. Yet their economic arguments should not be dismissed out of hand. Critics insist that much of the optimism about Turkey's economic roadmap has been over-egged - an argument amplified by a 134% rise in the country's current account deficit to \$ 10.7bn during the first 10 months of this year. The country's massive debt - which includes \$ 25bn owed to the IMF and billions borrowed via the international bond markets - also remains a major obstacle to its ambition of joining the EU. "In the new member states of the European Union, gross public debt is typically about 40% of gross domestic product," says Reza Moughadam, assistant director of the IMF's European Department. "At about 80% of GDP, Turkey's gross debt is double that figure." Turkey's debts have largely arisen from its efforts to push through banking reform after a run on the banks in 2001 caused the country's devastating recession. There is no question that although Turkey is doing much better than in the past, it remains quite vulnerable," says Michael Doppel, director of the IMF's European Department. "Its debt is far too high for an emerging economy. It should be reduced to a level that would allow Turkey to meet its obligations in a sustainable way. And although the state of Turkey's economy is important, as is its pace of reform, the final decision on 17 December will be taken by politicians who will, of course, be guided by their own interests."

TF-IDF Threshold=5.5

Turkey turns on the economic charm three years after a grueling economic crisis. Turkey has dressed its economy to impress. As part of a charm offensive - ahead of 17 December, when the European Union will decide whether to start entry talks - Turkey's economic leaders have been banging the drum to draw attention to recent achievements. The economy is growing fast, they insist. Education levels among its young and large population are rising. Unemployment levels, in percentage terms, are heading fast towards single digits. Inflation is under control. A new law to govern its turbulent banking system is on the cards. The tourism industry is booming and revenues from visitors should more than double to \$ 21bn (£10.8bn) in three years. Moreover, government spending is set to be frozen and a burdensome social security deficit is being tackled. Income and corporate taxes will be cut next year in order to attract \$ 15bn of foreign investment over the next three years. A loan restructuring deal with the International Monetary Fund (IMF) is pretty much in the can. And following recent macroeconomic restructuring efforts, its currency is floating freely and its central bank is independent. The point of all this has been to convince Europe's decision makers that rather than being a phenomenally costly exercise for the EU, allowing Turkey in would in fact bring masses of economic benefits. "The cake will be bigger for everybody," said Deputy Prime Minister Abdullahi Sener earlier this month. "Turkey will not be a burden for the EU budget." If admitted into the EU, Turkey would contribute almost 60n euros (\$ 8bn; £8n) to its budget by 2014, according to a recent impact study by the country's State Planning Organisation. As Turkey's gross domestic output (GDP) is set to grow by 6% per year on average, its contribution would rise from less than 5bn euros in 2014 to almost 20n euros by 2020. Turkey could also help alleviate a labour shortage in Old Europe once its population comes of age. By 2014, one in four Turks - of about 18 million people - will be aged 14 or less. "A literate and qualified Turkish population," insisted Mr Sener, "will make a positive impact on the EU." This runs contrary to the popular view that Turkey is getting ready to dip deep into EU taxpayers' wallets. However, Turkey's assertions are confirmed by Brussels' own impact studies, which indeed say that Turkish membership would be good news for the EU economy. But only over time. Costs are projected to be vast during the early years of Turkey's membership, with subsidies alone estimated to exceed 16.7bn euros and, according to some predictions, balloon to 33.7bn euros. This would include vast agricultural subsidies and regional aid; though such payments should decline as the country's farm sector, which currently employs one in three Turks, would employ just one in five by 2020. Such high initial expenses would be coupled with risks that the benefits flattered up by Turkey's government would never be delivered, say those who feel the Turkish project should be shunned. Some fear that rather than providing an educated, sophisticated labour force for Europe at large, the people who will leave Turkey to seek work abroad will be poor, uneducated and plentiful. More recently, less palatable concerns - at least in liberal European circles - have been voiced, with senior EU or member state officials talking darkly of a "river of Islam", an "oriental" culture and a threat to Europe's cultural richness. Of course, many opponents are politically motivated - their views ranging from xenophobic prejudices about the country's Muslim traditions to well-documented concerns about the government's human rights record. Yet their economic arguments should not be dismissed out of hand. Critics insist that much of the optimism about Turkey's economic roadmap has been over-egged - an argument amplified by a 134% rise in the country's current account deficit to \$ 10.7bn during the first 10 months of this year. The country's massive debt - which includes \$ 25bn owed to the IMF and billions borrowed via the international bond markets - also remains a major obstacle to its ambition of joining the EU. "In the new member states of the European Union, gross public debt is typically about 40% of gross domestic product," says Reza Moughadam, assistant director of the IMF's European Department. "At about 80% of GDP, Turkey's gross debt is double that figure." Turkey's debts have largely arisen from its efforts to push through banking reform after a run on the banks in 2001 caused the country's devastating recession. There is no question that although Turkey is doing much better than in the past, it remains quite vulnerable," says Michael Doppel, director of the IMF's European Department. "Its debt is far too high for an emerging economy. It should be reduced to a level that would allow Turkey to meet its obligations in a sustainable way. And although the state of Turkey's economy is important, as is its pace of reform, the final decision on 17 December will be taken by politicians who will, of course, be guided by their own interests."

TF-IDF Threshold=20

TF-IDF Threshold=50

ВИСНОВКИ

- Проаналізовані існуючі методи представлення знань та алгоритми пошуку та вилучення інформації з електронних текстів
- Розроблено алгоритм та програмний комплекс для побудови семантичних мереж знань, який можна застосовувати для автоматичного вилучення найбільш релевантних знань у неструктурованих текстах та побудови семантичних мереж знань
- В рамках дослідження була досягнута початкова мета та виконані поставлені задачі

ДЯКУЮ ЗА УВАГУ!

Студент групи ТІ-61 м
Савченко Микита Миколайович

Керівник
Крячок Олександр Степанович