МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ**

**«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

Інститут Комп’ютерних наук та інформаційних технологій

Кафедра Програмної інженерії та інтелектуальних технологій управління

Спеціальність 122 Комп’ютерні науки

Освітня програма Комп’ютерні науки та інтелектуальні системи

До захисту допускаю

Завідувач кафедри

Ігор ГАМАЮН

(ініціали та прізвище)

(підпис, дата)

**ДИПЛОМНА РОБОТА**

\_\_першого (бакалаврського)\_ рівня вищої освіти

Тема роботи ???

Шифр роботи КН-???.?

(група, номер теми за наказом)

Виконавець Рудський Олександр Вадимович

(прізвище, ім’я, по-батькові)

Керівник доцент Копп Андрій Михайлович (посада, прізвище, ім’я, по-батькові)

Харків 20??

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Найменування виробу,  об’єкту або теми | | | | Найменування  документу | | Фор-мат | | | | Кільк.  арк. | | При-  мітка |
|  | | | |  | |  | | | |  | |  |
|  | | | | Документи загальні | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | | Завдання на ДР | | А4 | | | | 1 | |  |
|  | | | | Пояснювальна записка до ДР | | А4 | | | | ?? | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | | Ілюстративні матеріали | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
| ??? | | | | Презентація | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | |  | | А4 | | | | ?? | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
|  | | | |  | |  | | | |  | |  |
|  |  |  |  | КН-???.? ВД | | | | | | | | |
|  |  |  |  |
|  | Прізвище | Підп | Дата |
| Розроб. | ??? |  |  | ???  Відомість документів | Літ. | | | | Аркуш | | Аркушів | |
| Перев. | КОПП |  |  | Д | | Р | Б |  | | 1 | |
|  |  |  |  | НТУ «ХПІ»  Кафедра „ПІІТУ” | | | | | | | |
| Н.конт. | ІВАЩЕНКО |  |  |
| Затв. | ГАМАЮН |  |  |

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ**

**«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

Інститут Комп’ютерних наук та інформаційних технологій

Кафедра Програмної інженерії та інтелектуальних технологій управління

Рівень вищої освіти перший (бакалаврський) рівень

Спеціальність 122 Комп’ютерні науки

Освітня програма Комп’ютерні науки та інтелектуальні системи

**ЗАТВЕРДЖУЮ**

**Завідувач кафедри**

Ігор ГАМАЮН

«\_\_\_\_» \_\_\_\_\_\_\_\_\_\_ 20\_\_ року

**З А В Д А Н Н Я**

**НА ДИПЛОМНУ РОБОТУ СТУДЕНТУ**

???

(прізвище, ім’я, по батькові)

1 Тема роботи ???

керівник роботи Копп Андрій Михайлович, доктор філософії

(прізвище, ім’я, по батькові, науковий ступінь, вчене звання)

затверджені наказом вищого навчального закладу від «??» ??? 20?? року № ??? СТ

2 Строк подання студентом роботи «??» ??? 20?? року

3 Вихідні дані до роботи Літературні джерела. Ресурси мережі Інтернет.

4 Перелік питань, які потрібно розробити у пояснювальній записці

???

5 Перелік графічного матеріалу (з точним зазначенням обов’язкових креслень)

Рисунків – ??, таблиць – ??.

6 Консультанти розділів роботи

|  |  |  |  |
| --- | --- | --- | --- |
| Розділ | Прізвище, ініціали та посада  консультанта | Підпис, дата | |
| завдання  видав | завдання  прийняв |
|  |  |  |  |
|  |  |  |  |

7 Дата видачі завдання «??» ??? 20?? року

**КАЛЕНДАРНИЙ ПЛАН**

|  |  |  |  |
| --- | --- | --- | --- |
| Номер етапу | Назва етапів дипломної роботи | Строк  виконання  етапів роботи | Примітка |
| 1 | Вибір і затвердження теми випускної роботи |  |  |
| 2 | Розробка і затвердження календарного плану роботи |  |  |
| 3 | Огляд методів щодо вирішення задачі ??? |  |  |
| 4 |  |  |  |
| 5 |  |  |  |
| 6 | Оформлення звіту |  |  |
| 7 | Розробка демонстраційних плакатів |  |  |
| 8 | Складання відомості документів (ВД) |  |  |
| 9 | Оформлення та комплектування ДР |  |  |
| 10 | Подання закінченої ДР на допуск до захисту |  |  |
| 11 | Рецензування дипломної роботи |  |  |
| 12 | Захист ДР |  |  |

**Студент**  ???

(підпис) (прізвище та ініціали)

**Керівник роботи**  Андрій КОПП

(підпис) (прізвище та ініціали)

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ**

**«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

Інститут Комп’ютерних наук та інформаційних технологій

Кафедра Програмної інженерії та інтелектуальних технологій управління

Спеціальність 122 Комп’ютерні науки

Освітня програма Комп’ютерні науки та інтелектуальні системи

**ПОЯСНЮВАЛЬНА ЗАПИСКА**

**до дипломної роботи**

\_\_\_\_\_\_першого (бакалаврського)\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ рівня вищої освіти

на тему \_\_???\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Виконав студент 4 курсу, групи КН-419б

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_??? \_\_\_\_\_\_\_\_\_\_

(підпис, прізвище та ініціали)

Керівник \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Андрій КОПП

(підпис, прізвище та ініціали)

Рецензентка \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_???

(підпис, прізвище та ініціали)

Нормоконтроль \_\_\_\_\_\_\_\_Оксана ІВАЩЕНКО

(підпис, прізвище та ініціали)

Харків 20??

**РЕФЕРАТ**

Пояснювальна записка до ДР: ?? с., ?? рис., ?? табл., ?? джерел

*Ключові слова*: ТЕКСТ.

Об’єкт роботи – процес аналізу відповідності моделей бізнес-процесів їх текстовим описам.

Предмет роботи – алгоритмічне та програмне забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам.

Мета роботи – забезпечення адекватності моделей бізнес-процесів за рахунок аналізу їх відповідності текстовим описам.

Таким чином, для досягнення мети, у роботі було вирішено наступні задачі:

* ???.

**ABSTRACT**

Report of implementation of DW: ?? p., ?? fig., ?? tab., ?? sources

*Keywords*: TEXT.

The object of work is ???.

The subject of work is ???.

The purpose of the work is ???.

Thus, to achieve the goal, the following tasks were solved in the work:

* ???.

ЗМІСТ

[Перелік позначень та скорочень 4](#_Toc118632035)

[Вступ 5](#_Toc118632036)

[1 Аналіз існуючих рішень щодо моделювання бізнес-процесів та формування вимог до програмного забезпечення 6](#_Toc118632037)

[1.1 Аналіз існуючих інструментальних засобів моделювання та аналізу бізнес-процесів 6](#_Toc118632038)

[1.2 Виявлення та специфікація вимог до програмного забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам 9](#_Toc118632039)

[1.2.1 Виявлення вимог до програмного забезпечення 9](#_Toc118632040)

[1.2.2 Специфікація вимог до програмного забезпечення 11](#_Toc118632041)

[1.3 Мета та задачі роботи 12](#_Toc118632042)

[2 Алгоритмічне забезпечення для розв’язання задачі аналізу відповідності моделей бізнес-процесів їх текстовим описам 14](#_Toc118632043)

[2.1 Розробка алгоритмічного забезпечення для розв’язання задачі аналізу відповідності моделей бізнес-процесів їх текстовим описам 14](#_Toc118632044)

[2.2 Моделювання процесу аналізу відповідності моделей бізнес-процесів їх текстовим описам 18](#_Toc118632045)

[3 Проєктування та розробка програмного забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам 19](#_Toc118632046)

[3.1 Проєктування архітектури програмного забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам 19](#_Toc118632047)

[3.1.1 Вибір типу архітектури програмного забезпечення 19](#_Toc118632048)

[3.1.2 Моделювання архітектури програмного забезпечення 19](#_Toc118632049)

[3.2 Проєктування та розробка бази даних програмного забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам 20](#_Toc118632050)

[3.2.1 Вибір системи управління базами даних 20](#_Toc118632051)

[3.2.2 Моделювання та реалізація бази даних 20](#_Toc118632052)

[3.3 Проєктування та розробка програмного забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам 21](#_Toc118632053)

[4 Застосування програмного забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам та аналіз отриманих результатів 22](#_Toc118632054)

[4.1 Застосування програмного забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам 22](#_Toc118632055)

[4.2 Аналіз отриманих результатів щодо перевірки відповідності моделей бізнес-процесів їх текстовим описам 22](#_Toc118632056)

[Висновки 23](#_Toc118632057)

[Список джерел інформації 24](#_Toc118632058)

# Перелік позначень та скорочень

СУБД – система управління базами даних.

BPM – Business Process Management.

BPMN – Business Process Model and Notation.

NLP – Natural Language Processing.

UML – Unified Modeling Language.

# Вступ

Моделі бізнес-процесів зарекомендували себе як ефективний засіб візуалізації та вдосконалення складних організаційних операцій. Моделі бізнес-процесів використовуються для пошуку неефективних місць в описаних бізнес-процесах та усунення виявлених недоліків шляхом автоматизації за допомогою програмних рішень, що налаштовуються, або уніфікованих програмних середовищ виконання бізнес-процесів [1].

BPM (Business Process Management, управління бізнес-процесами) – це концепція управління організацією на рівні процесів, які розглядаються як бізнес-ресурс, що постійно змінюється і адаптується до змін всередині та в бізнес-кліматі. Основні принципи цієї концепції – це прозорість та зрозумілість бізнес-процесів. Для досягнення цієї мети вдаються до моделювання процесів шляхом використання нотацій певного сталого стандарту. Найпоширенішим стандартом такої нотації є BPMN 2.0 [2] (Business Process Model and Notation, нотація моделювання бізнес-процесів). Нотація містить постійний перелік елементів, що застосовується для побудови діаграм – моделей бізнес-процесів. Така модель може включати в себе набір дій та подій.

Однак створення моделей бізнес-процесів є трудомісткою справою, яка потребує значних ресурсів, тому можуть виникати ситуації в яких модель бізнес-процесу не відповідає її текстовому опису. Це може призвести до втрат у часі та значних грошових втрат. Таким чином, актуальною є задача аналізу відповідності моделей бізнес-процесів їх текстовим описам [1].

Об’єкт роботи – процес аналізу відповідності моделей бізнес-процесів їх текстовим описам.

Предмет роботи – алгоритмічне та програмне забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам.

Мета роботи – забезпечення адекватності моделей бізнес-процесів за рахунок аналізу їх відповідності текстовим описам.

# 1 Аналіз існуючих рішень щодо моделювання бізнес-процесів та формування вимог до програмного забезпечення

## 1.1 Аналіз існуючих інструментальних засобів моделювання та аналізу бізнес-процесів

Першим серед розглянутих аналогів є Bonita BPM. Bonita BPM – це програмне забезпечення для управління бізнес-процесами. Це програма, яка встановлюється на комп’ютер аналітика або розробника. За допомогою цього інструменту можна редагувати BPMN-схеми, створювати моделі даних, завантажувати довідники користувачів, малювати форми [3].

Інтерфейс Bonita BPM наведено на рисунку 1.1.

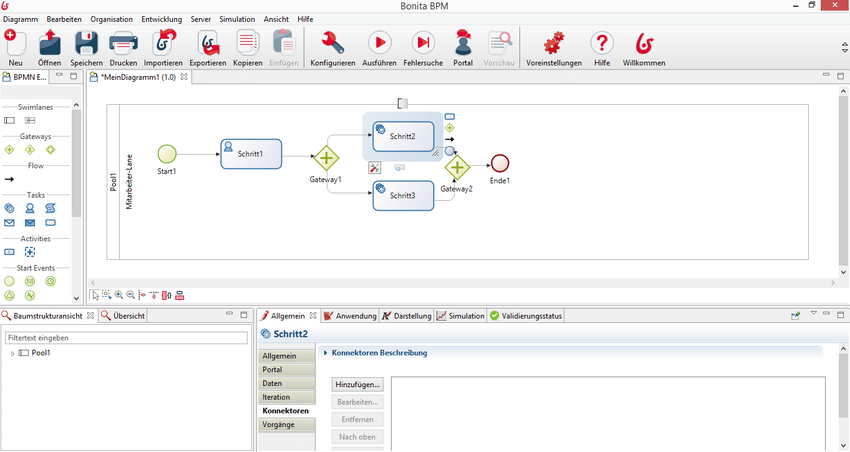


Рисунок 1.1 – Інтерфейс Bonita BPM

Переваги Bonita BPM:

* відносно простий у користуванні;

Недоліки Bonita BPM:

* потрібно завантажувати застосунок;
* неможливо проаналізувати відповідність моделі бізнес-процесів її текстовому опису.

Наступним серед аналогів є інший додаток – Signavio. Signavio Process Manager – це веб-рішення для проектування, аналізу (моделювання) та документування бізнес-процесів. Це рішення дозволяє створювати моделі процесів у стилі блок-схем безпосередньо у браузері, зв’язувати будь-який документ із процесами (робочі процедури, регламенти, вказівки щодо надання послуг тощо), документувати рішення в рамках процесів у графічному вигляді, експортувати процеси у різних форматах (.png, .svg, .pdf, xml для BPMN 2.0) [4].

Інтерфейс Signavio Process Manager наведено на рисунку 1.2.

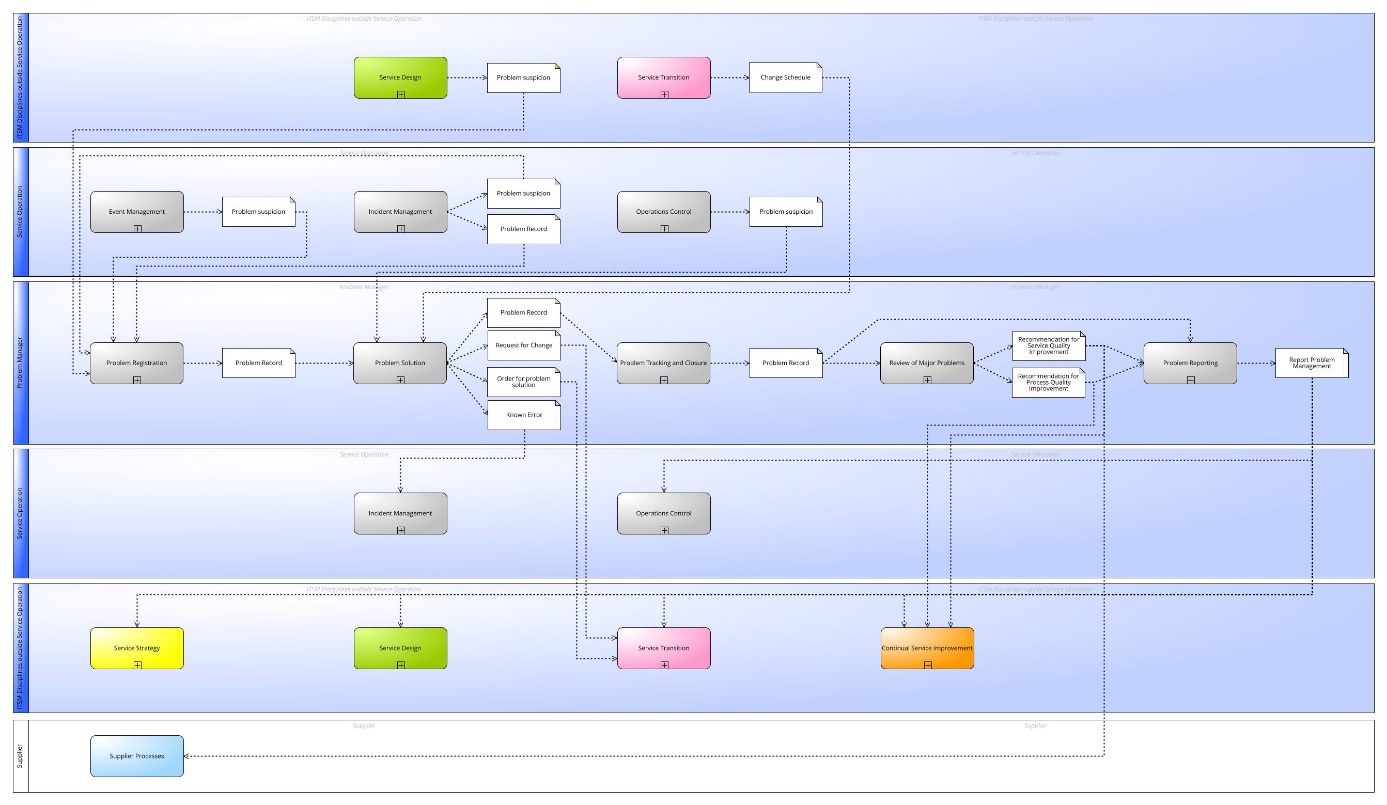


Рисунок 1.2 – Інтерфейс Signavio Process Manager

Переваги Signavio Process Manager:

* відносно простий у користуванні;
* не потрібно завантажувати застосунок, так як це веб-додаток;

Недоліки Signavio Process Manager:

* неможливо проаналізувати відповідність моделі бізнес-процесів її текстовому опису.

Ще один із аналогів – ProcessMaker – це програмне забезпечення для управління бізнес-процесами (BPM). Воно дозволяє користувачам ефективно моделювати свої бізнес-процеси.

Програмне забезпечення повністю доступне через Інтернет та доступне через будь-який веб-браузер, що спрощує управління та координацію бізнес-процесів у всій організації [5].

Інтерфейс ProcessMaker наведено на рисунку 1.3.

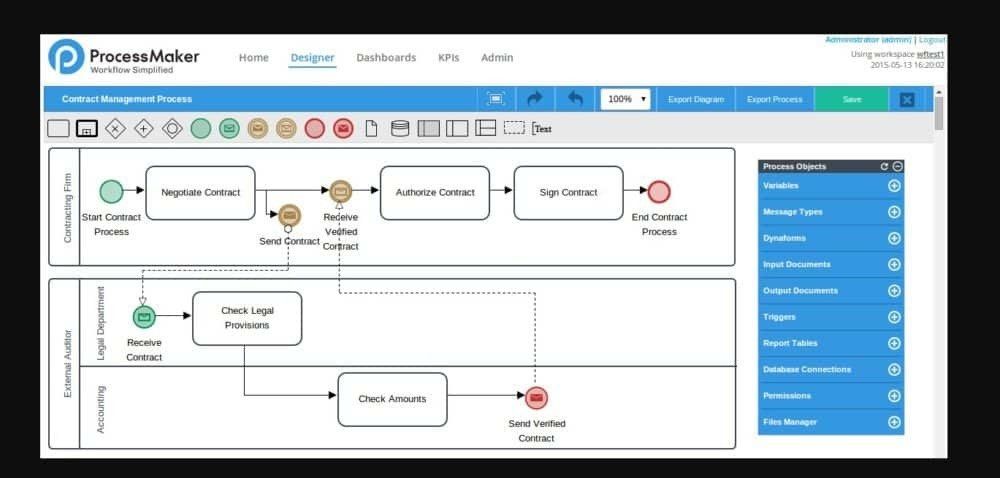


Рисунок 1.3 – Інтерфейс ProcessMaker

Переваги ProcessMaker:

* відносно простий у користуванні;
* не потрібно завантажувати застосунок, так як це веб-додаток;

Недоліки ProcessMaker:

* неможливо проаналізувати відповідність моделі бізнес-процесів її текстовому опису.

Bizagi – це програмне забезпечення для побудови карт та моделей бізнес-процесів у нотації BPMN. Воно дозволяє створювати, інтерпретувати та оптимізувати діаграми робочих процесів з використанням нотації BPMN, публікувати документацію щодо бізнес-процесу у форматах Word, PDF, Excel, Wiki [6].

Інтерфейс Bizagi наведено на рисунку 1.4.

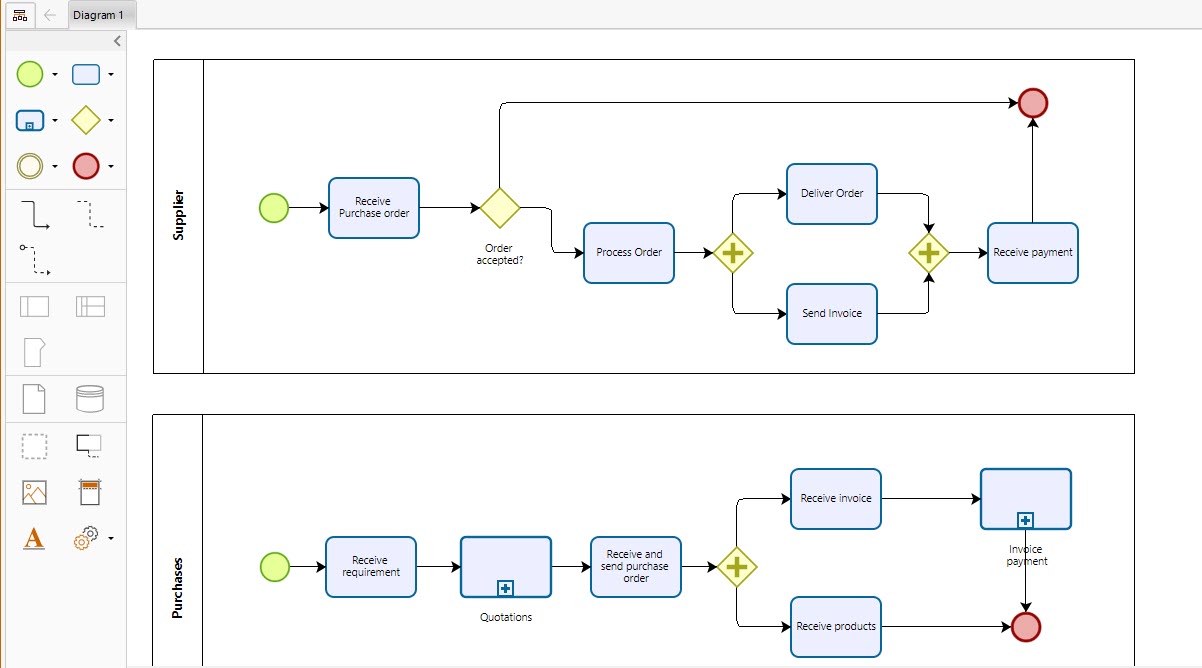


Рисунок 1.4 – Інтерфейс Bizagi

Переваги Bizagi:

* відносно простий у користуванні;

Недоліки Bizagi:

* потрібно завантажувати застосунок;
* неможливо проаналізувати відповідність моделі бізнес-процесів її текстовому опису.

## 1.2 Виявлення та специфікація вимог до програмного забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам

### 1.2.1 Виявлення вимог до програмного забезпечення

Після проведеного аналізу існуючих програмних засобів можна зробити висновок, що всі застосунки значно спрощують процес побудови моделей бізнес-процесів, але ці застосунки не надають можливості аналізувати моделі на наявність помилок. Насамперед – у них відсутня можливість порівнювати моделі бізнес-процесів з їхніми текстовими описами, що може привести до втрат у часі та значних грошових втрат при реалізації неадекватних предметній області моделей.

Функціональні вимоги – це вимоги до програмного забезпечення, які описують внутрішню роботу системи, її поведінку: обчислення даних, маніпулювання даними, обробка даних та інші специфічні функції, які має виконувати система. На відміну від нефункціональних вимог, які визначають якою система повинна бути, функціональні вимоги визначають, що система повинна робити [7].

Для програмного рішення було визначено наступні функціональні вимоги:

* програмне забезпечення має обробляти завантажений файл BPMN;
* програмне рішення має порівнювати модель бізнес-процесу, отриману з файлу BPMN, із текстовим описом бізнес-процесу;
* застосунок повинен відображати результати у вигляді відповідного звіту;
* застосунок повинен мати можливість реєстрації для користувачів;
* програмне рішення повинне зберігати звіт з результатами у базі даних для зареєстрованих користувачів.

Нефункціональні вимоги – це вимоги до програмного забезпечення, які задають критерії для оцінки якості його роботи. На відміну від функціональних вимог, які визначають що система повинна робити, нефункціональні вимоги визначають якою система повинна бути. Нефункціональні вимоги до програмного забезпечення визначаються на першій стадії процесу розробки ПЗ – на етапі аналізу вимог [7].

Для програмного рішення було визначено наступні нефункціональні вимоги:

* архітектура – програмне забезпечення повинно бути веб-додатком і мати трирівневу архітектуру;
* не перевантажений інтерфейс – на екрані повинно бути не більше п’яти елементів управління (кнопок, полів вводу тощо);
* тип системи управління базами даних (СУБД) – програмне забезпечення повинно використовувати реляційну СУБД;
* надійність – програмне забезпечення не повинно містити помилок, які можуть перешкоджати його роботі.

### 1.2.2 Специфікація вимог до програмного забезпечення

У відповідності до сформульованих функціональних вимог до програмного рішення було побудовано UML-діаграму варіантів використання, яку наведено на рисунку 1.5.

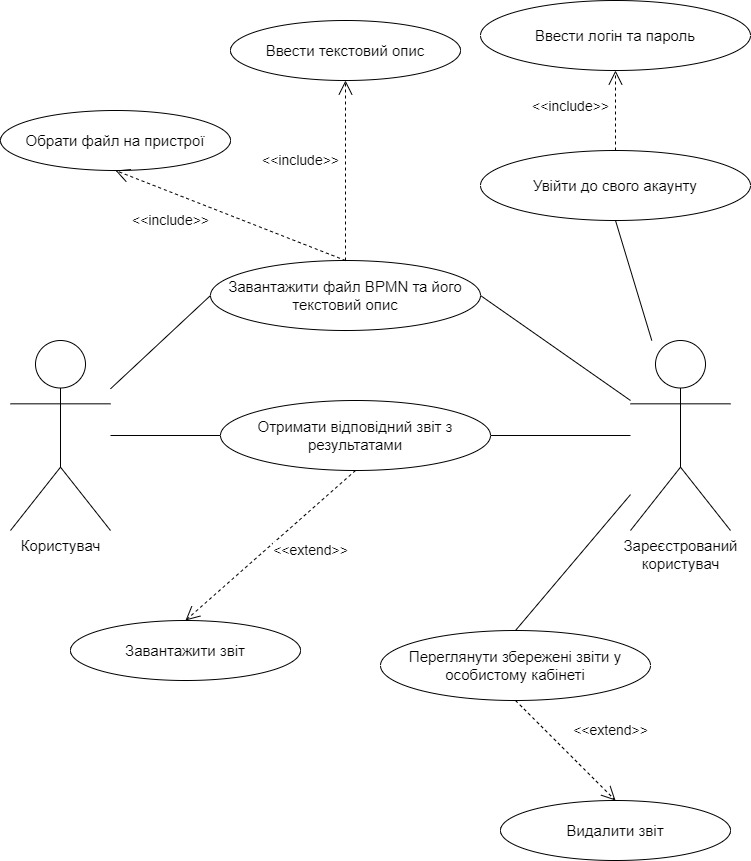


Рисунок 1.5 – UML-діаграма варіантів використання

На рисунку 1.6 наведено діаграму нефункціональних вимог, яка визначає обмеження якості програмного рішення.

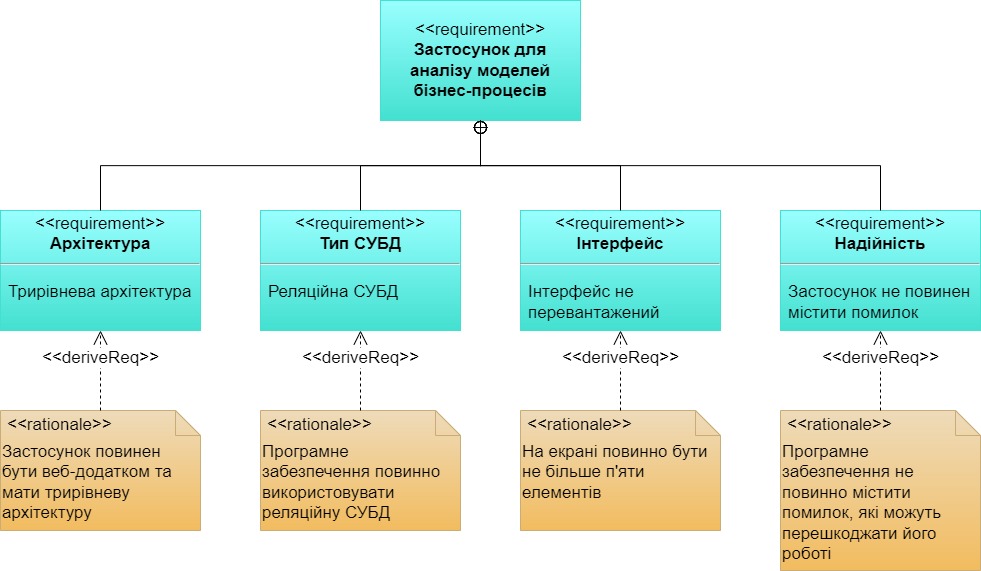


Рисунок 1.6 – Діаграма обмежень якості

Користувацькі історії (User Story) — спосіб опису вимог до системи, що розробляється, сформульованих як одна або більше пропозицій повсякденною або діловою мовою користувача. Історії користувача використовуються гнучкими методологіями розробки програмного забезпечення для специфікації вимог. Кожна користувальницька історія обмежена у розмірі та складності. Мета історій користувача полягає в тому, щоб бути в змозі оперативно і без накладних витрат реагувати на вимоги реального світу, що швидко змінюються [8].

Для програмного рішення було визначено наступні користувацькі історії:

* як користувач застосунку, я хочу порівнювати модель бізнес-процесів із їх текстовим описом;
* як користувач застосунку, я хочу мати можливість завантажити звіт з результатами;
* як користувач застосунку, я хочу мати можливість зареєструватися у додатку;
* як зареєстрований користувач, я хочу переглядати збережені звіти у своєму особистому кабінеті;
* як зареєстрований користувач, я хочу мати можливість видаляти збережені звіти.

## 1.3 Мета та задачі роботи

Існуючі на сьогоднішній день програмні засоби передбачають побудову моделей бізнес-процесів, але не надають можливості аналізувати ці моделі з точки зору їх адекватності реальним бізнес-процесам, а саме – порівнювати моделі бізнес-процесів з їх текстовим описом. Тому актуальною є розробка алгоритмічного та програмного забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам.

Об’єкт роботи – процес аналізу відповідності моделей бізнес-процесів їх текстовим описам.

Предмет роботи – алгоритмічне та програмне забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам.

Мета роботи – забезпечення адекватності моделей бізнес-процесів за рахунок аналізу їх відповідності текстовим описам.

Таким чином, для досягнення мети роботи необхідно:

* розробити алгоритмічне забезпечення розв’язання задачі аналізу відповідності моделей бізнес-процесів їх текстовим описам;
* проаналізувати процес перевірки моделей бізнес-процесів щодо відповідності їх текстовим описам;
* обрати тип архітектури програмного забезпечення, побудувати моделі архітектури програмного забезпечення, спроектувати та розробити базу даних;
* спроектувати та розробити програмне забезпечення;
* застосувати програмне забезпечення для розв’язання задачі аналізу відповідності моделей бізнес-процесів їх текстовим описам;
* провести аналіз отриманих результатів.

# 2 Алгоритмічне забезпечення для розв’язання задачі аналізу відповідності моделей бізнес-процесів їх текстовим описам

## 2.1 Розробка алгоритмічного забезпечення для розв’язання задачі аналізу відповідності моделей бізнес-процесів їх текстовим описам

Програмне забезпечення для розв’язання задачі аналізу відповідності моделей бізнес-процесів їх текстовим описам має вилучати з файлу BPMN усі назви елементів типу «task» та споріднених елементів дій (рисунок 2.1):

* Service Task;
* Send Task;
* Receive Task;
* User Task;
* Manual Task;
* Business Rule Task;
* Script Task;

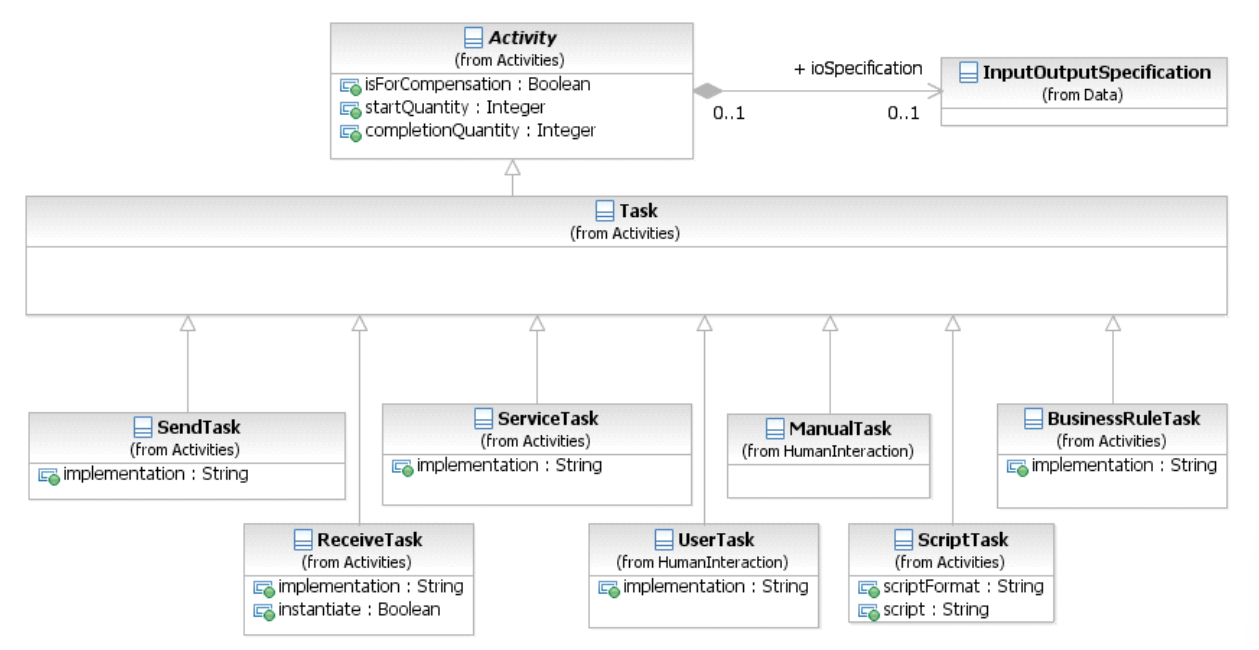


Рисунок 2.1 – Типи дій нотації BPMN [9]

Таким чином, необхідно сформувати текст з усіх назв елементів типу «task» та споріднених елементів дій (рисунок 2.1).

Для цього може використовуватись наступний алгоритм, представлений UML-діаграмою активності на рисунку 2.2.

Алгоритм формування тексту з назв тасків моделі БП

Рисунок 2.2 – UML-діаграма активності

На вхід разом з файлом BPMN моделі бізнес-процесу подається також і текстовий опис бізнес-процесу, який має представляти ця модель.

Для того, щоб обробляти дані тексти з метою аналізу відповідності моделей бізнес-процесів їх текстовим описам, в роботі пропонується застосувати наступні технології обробки природної мови або NLP-технології (Natural Language Processing):

* токенізація;
* пошук стоп-слів;
* стемінг.

Токенізація – це перший крок у будь-якому процесі NLP. Токенізатор розбиває неструктуровані дані та текст природною мовою на шматки інформації, які можна розглядати як дискретні елементи. Це відразу перетворює неструктурований рядок (текстовий документ) на числову структуру даних, придатну для машинного навчання. Вони також можуть використовуватися безпосередньо комп'ютером для запуску корисних дій та відповідей. Або вони можуть використовуватися в процесі машинного навчання як функції, які викликають більш складні рішення або поведінку [10].

Токенізація може бути виконана на рівні слова, символу або підмови [11]:

* токенізація слів є найчастіше використовуваним алгоритмом токенізації. Він розбиває шматок тексту окремі слова з урахуванням певного роздільника. Залежно від роздільників формуються різноманітні маркери рівня слів [11];
* токенізація символів розбиває частину тексту на набір символів [11];
* Токенізація підслів розбиває фрагмент тексту на підслова (або n-грамні символи). Наприклад, такі слова як “lower” можна сегментувати як “low-er”, “smartest” як “smart-est” і так далі [11].

Після того, як ми розділили текст на токени, часто стає ясно, що не всі слова несуть однакову кількість інформації, якщо взагалі будь-яку інформацію для завдання прогнозного моделювання. Загальні слова, які несуть мало значної інформації, називаються стоп-словами. Стоп-слова – це слова будь-якою мовою, які не додають великого сенсу пропозиції. Їх можна сміливо ігнорувати, не жертвуючи змістом речення. Для деяких пошукових систем це деякі з найпоширеніших, коротких функціональних слів, таких як “the”, “is”, “at”, “which” та “on”. У цьому випадку стоп-слова можуть викликати проблеми при пошуку фраз, які їх включають[12].

Якщо ми маємо завдання класифікації тексту чи аналізу тональності, ми повинні видалити стоп-слова, оскільки вони не надають жодної інформації для нашої моделі, тобто виключають небажані слова з нашого корпусу, але якщо у нас є завдання мовного перекладу, то стоп-слова корисні, оскільки вони мають бути перекладені разом з іншими словами[12].

Стемінг є однією з найпоширеніших операцій попередньої обробки даних, яку ми виконуємо майже у всіх проектах обробки природної мови (NLP). Стемінг – це процес скорочення слова до основи шляхом відкидання допоміжних частин, таких як закінчення чи суфікс. Результати стемінгу іноді дуже схожі на визначення кореня слова, але його алгоритми базуються на інших принципах. Тому слово після обробки алгоритмом стемінгу може відрізнятися від морфологічного кореня слова [13].

Існує кілька варіантів алгоритмів стемінгу, які відрізняються своєю точністю та продуктивністю:

* пошук за таблицею;
* відсічення закінчень та суфіксів;
* лематизація;
* стохастичні алгоритми;
* гібридний підхід;
* відсічення префіксів;
* пошук відповідності.

Таким чином, для розв’язання задачі аналізу відповідності моделей бізнес-процесів їх текстовим описам з отриманими текстами та необхідно виконати наступні дії:

* розбити отримані на вхід тексти та на окремі слова (токенізувати), отримавши відповідні мультимножини слів:

(2.1)

де – множина слів, отриманих в результаті токенізації тексту ;  
 – множина слів, отриманих в результаті токенізації тексту ;  
, – слово, отримане в результаті токенізації тексту ;  
, – слово, отримане в результаті токенізації тексту ;  
 – відображення , яке для кожного слова , встановлює кількість його повторювань в тексті ;  
 – відображення , яке для кожного слова , встановлює кількість його повторювань в тексті ;  
 – кількість слів, отриманих в результаті токенізації тексту ;  
 – кількість слів, отриманих в результаті токенізації тексту ;

* видалити стоп-слова з множин та , отримавши множини лише змістовних термінів, що стосуються предметної області бізнес-процесу:

(2.2)

де , – множина слів, отримана в результаті токенізації вихідного тексту, що також містить стоп-слова;  
, – множина слів, отримана в результаті токенізації вихідного тексту, з якої були видалені стоп-слова;  
 – відображення, яке для кожної множини , , яка містить стоп-слова, ставить у відповідність множину , , яка не містить стоп-слова;  
 – кількість множин слів, що обробляються, ;

* виконати стемінг слів у множинах та , що залишились після видалення стоп-слів:

(2.3)

де , – множина слів, отримана в результаті стемінгу слів, що залишились після видалення стоп-слів;  
 – відображення, яке для кожної множини , , з якої були видалені стоп-слова, ставить у відповідність множину , , слова в якій, що залишились після видалення стоп-слів, були скорочені до основи.

Таким чином, в результаті виконання попередній дій, будуть отримані дві множини слів та :

(2.4)

Обчислити подібність цих двох множин слів та можна за допомогою однієї з наступних метрик [??]:

* коефіцієнт Жаккара:

; (2.5)

* коефіцієнт Соренсена:

формула; (2.6)

* коефіцієнт Кульчинскього:

формула; (2.7)

* коефіцієнт Сімпсона:

формула; (2.8)

* коефіцієнт Браун-Бланке:

формула; (2.9)

* коефіцієнт Отіаі:

формула. (2.10)

обґрунтувати вибір одного з цих коефіцієнтів для використання в алгоритмі розв’язання задачі

Відповідно, отримане значення коефіцієнту Жаккра можна інтерпретувати як ступінь відповідності моделі бізнес-процесу її текстовому опису.

Таким чином, у роботі пропонується наступний алгоритм розв’язання задачі аналізу відповідності моделей бізнес-процесів їх текстовим описам, представлений UML-діаграмою активності на рисунку 2.3.

Детальна схема алгоритму починаючи з токенізації текстів і закінчуючи розрахунком коефіцієнту схожості

Рисунок 2.3 – Алгоритм розв’язання задачі аналізу відповідності моделей бізнес-процесів їх текстовим описам

## 2.2 Моделювання процесу аналізу відповідності моделей бізнес-процесів їх текстовим описам

Контекстна діаграма (IDEF0) потоку робіт.

Рисунок 2.2 – Контекстна діаграма потоку робіт

Діаграма декомпозиції (IDEF0).

Рисунок 2.3 – Діаграма декомпозиції потоку робіт

???

# 3 Проєктування та розробка програмного забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам

## 3.1 Проєктування архітектури програмного забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам

### 3.1.1 Вибір типу архітектури програмного забезпечення

Короткий огляд та опис архітектурних стилів.

Рисунок 3.1 – ???

???

Рисунок 3.2 – ???

???

Рисунок 3.3 – ???

Обґрунтований вибір архітектури програмного забезпечення.

### 3.1.2 Моделювання архітектури програмного забезпечення

Обґрунтований вибір програмної платформи та технологічного стеку.

UML діаграма розгортання.

Рисунок 3.4 – UML-діаграма розгортання

UML діаграма компонентів.

Рисунок 3.5 – UML-діаграма компонентів

???

## 3.2 Проєктування та розробка бази даних програмного забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам

### 3.2.1 Вибір системи управління базами даних

Короткий огляд та опис СУБД.

Рисунок 3.6 – ???

Обґрунтований вибір СУБД для реалізації бази даних.

### 3.2.2 Моделювання та реалізація бази даних

Контекстна модель даних (ER діаграма).

Рисунок 3.7 – Логічна модель даних

Логічна модель даних (IDEF1X діаграма).

Рисунок 3.8 – Фізична модель даних

???

## 3.3 Проєктування та розробка програмного забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам

Обґрунтований вибір інструментів розробки.

UML діаграма класів (структурована за пакетами = компонентами).

Рисунок 3.9 – UML-діаграма класів

UML діаграма послідовності.

Рисунок 3.10 – UML-діаграма послідовності

UML діаграма діяльності (зображує бізнес-процес з точки зору програмного забезпечення).

Рисунок 3.11 – UML-діаграма діяльності

???

# 4 Застосування програмного забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам та аналіз отриманих результатів

## 4.1 Застосування програмного забезпечення для аналізу відповідності моделей бізнес-процесів їх текстовим описам

Аналіз відповідності програмного забезпечення користувацьким історіям (зі скріншотами інтерфейсу користувача).

Рисунок 4.1 – ???

???

## 4.2 Аналіз отриманих результатів щодо перевірки відповідності моделей бізнес-процесів їх текстовим описам

Оцінити отримані результати, наприклад, за допомогою крос-валідації (cross-validation).

x + y = z, (4.1)

де x – ???;  
y – ???.

a + b = c. (4.2)

???

# Висновки

Короткий опис того, що було виконано у кожному розділі.

Результати вирішення задач (сформульованих у 1.3).

Подальший розвиток роботи (що може бути зроблено в майбутньому для покращення отриманих результатів).

# Список джерел інформації

1. Overview of Verification Tools for Business Process Models // URL: https://www.researchgate.net/publication/320012697\_Overview\_of\_Verification\_Tools\_for\_Business\_Process\_Models/, 25.10.2022.
2. Thomas Allweyer. BPMN 2.0: Introduction to the Standard for Business Process Modeling – 174 сторінки
3. Be Efficient: Bonitasoft Introduces New Bonita BPM 6 Platform // URL: https://www.businesswire.com/news/home/20130605006087/en/Efficient-Bonitasoft-Introduces-Bonita-BPM-6-Platform#.VNM5HC5eT5E/, 25.10.2022.
4. SAP Signavio // URL: https://www.signavio.com/, 25.10.2022.
5. ProcessMaker Review // URL: https://comparecamp.com/processmaker-review-pricing-pros-cons-features/, 25.10.2022.
6. Process Documentation - Bizagi Models // URL: https://help.bizagi.com/process-modeler/en/index.html?where\_to\_share2.htm, 25.10.2022.
7. Функціональні та Не Функціональні Вимоги // URL: http://lvivqaclub.blogspot.com/2008/10/blog-post\_17.html/, 25.10.2022.
8. Cohn, Mike. User Stories Applied: For Agile Software Development. Addison-Wesley, 2004 – 304 c.
9. Business Process Model and Notation (BPMN) // URL: https://www.omg.org/spec/BPMN/2.0/PDF, 10.11.2022.
10. Tokenization in NLP: Types, Challenges, Examples, Tools // URL: https://neptune.ai/blog/tokenization-in-nlp, 10.11.2022.
11. What is Tokenization | Tokenization In NLP // URL: https://www.analyticsvidhya.com/blog/2020/05/what-is-tokenization-nlp/, 10.11.2022.
12. Stop Words in NLP // URL: https://medium.com/@saitejaponugoti/stop-words-in-nlp-5b248dadad47, 10.11.2022.
13. Jongejan, B. and H. Dalianis. Automatic training of lemmatization rules that handle morphological changes in pre-, in- and suffixes alike, 2009 – 153 с.

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ**

**«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

Інститут Комп’ютерних наук та інформаційних технологій

Кафедра Програмної інженерії та інтелектуальних технологій управління

Спеціальність 122 Комп’ютерні науки

Освітня програма Комп’ютерні науки та інтелектуальні системи\_\_\_\_\_\_\_\_

**ІЛЮСТРАТИВНІ МАТЕРІАЛИ**

Тема роботи  ???

Шифр роботи  КН-???.?

(група, номер теми за наказом)

Виконавець                ???

(прізвище, ім’я, по-батькові)

Керівник доцент Андрій КОПП

(посада, прізвище, ім’я, по-батькові)

Харків 20??