Програмне забезпечення для зберігання та обробки неструктурованих документів

Виконала студентка
КПІ ім. Ігоря Сікорського, ТЕФ,
Групи ТР-71мп
Масечко І. О.

Керівник: к.т.н. Михайлова І. Ю.

Проблема

- Значна частина усієї інформації, яка використовується підприємствами для щоденної роботи, зберігається у вигляді паперових документів або їх відсканованих копій. Це особливо стосується даних, отриманих із зовнішніх джерел. Тобто більшість вхідних даних компаній є неструктурованими.
- **Неструктуровані дані** – інформація, яка не має попередньої певної структури даних, або не організована в установленому порядку. Це призводить до труднощів аналізу, особливо у випадку використання традиційних програм, призначених для роботи зі структурованими даними.

Мета та основні завдання дисертації

Метою магістерської дисертації є розробка програмної системи для зберігання та обробки неструктурованих документів у мультимодельній СКБД Caché.

Для досягнення поставленої мети необхідно вирішити завдання:

— проаналізувати існуючі програмні системи для роботи з неструктурованими документами, визначити їх переваги та недоліки;

— обрати засоби реалізації програмної системи для роботи з неструктурованими документами;

— розробити архітектуру та структуру програмної системи для роботи з неструктурованими документами;

— розробити програмну систему для роботи з неструктурованими документами.

Програмна система для роботи з неструктурованими документами повинна бути розроблена з використанням СКБД InterSystems Caché та технології InterSystems iKnow.

Мета та основні завдання системи

Основною метою розробленої програми для управління неструктурованими документами є забезпечення обліково-реєстраційних, класифікаційних та інформаційно-пошукових функцій.

Програмна реалізація системи

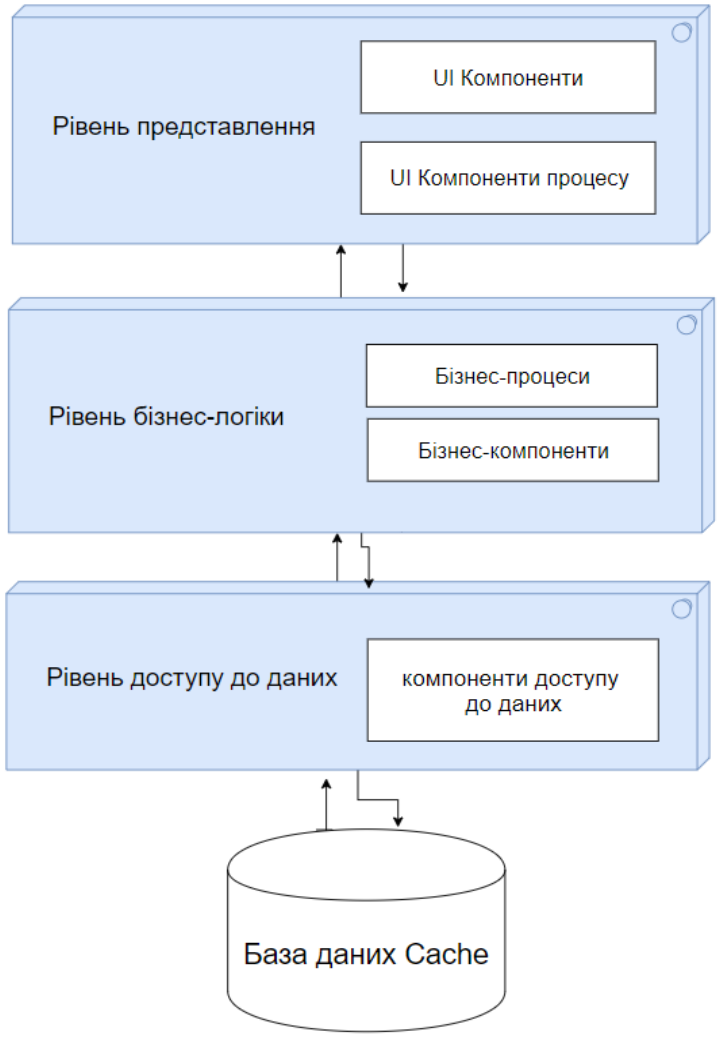
Для ефективної реалізації зазначеної задачі було вирішено використати постреляційну базу даних InterSystems Caché та технологію роботи з неструктурованими даними iKnow для автоматичного визначення тегів текстових документів.

Система керування базами даних (СКБД) InterSystems Caché дозволяє швидко обробляти дані при реалізації складних систем, використовувати об'єктно-орієнтований підхід до проектування, створювати класи з властивостями різних типів даних та об'єкти, до яких, за необхідності, можна отримати реляційний доступ .

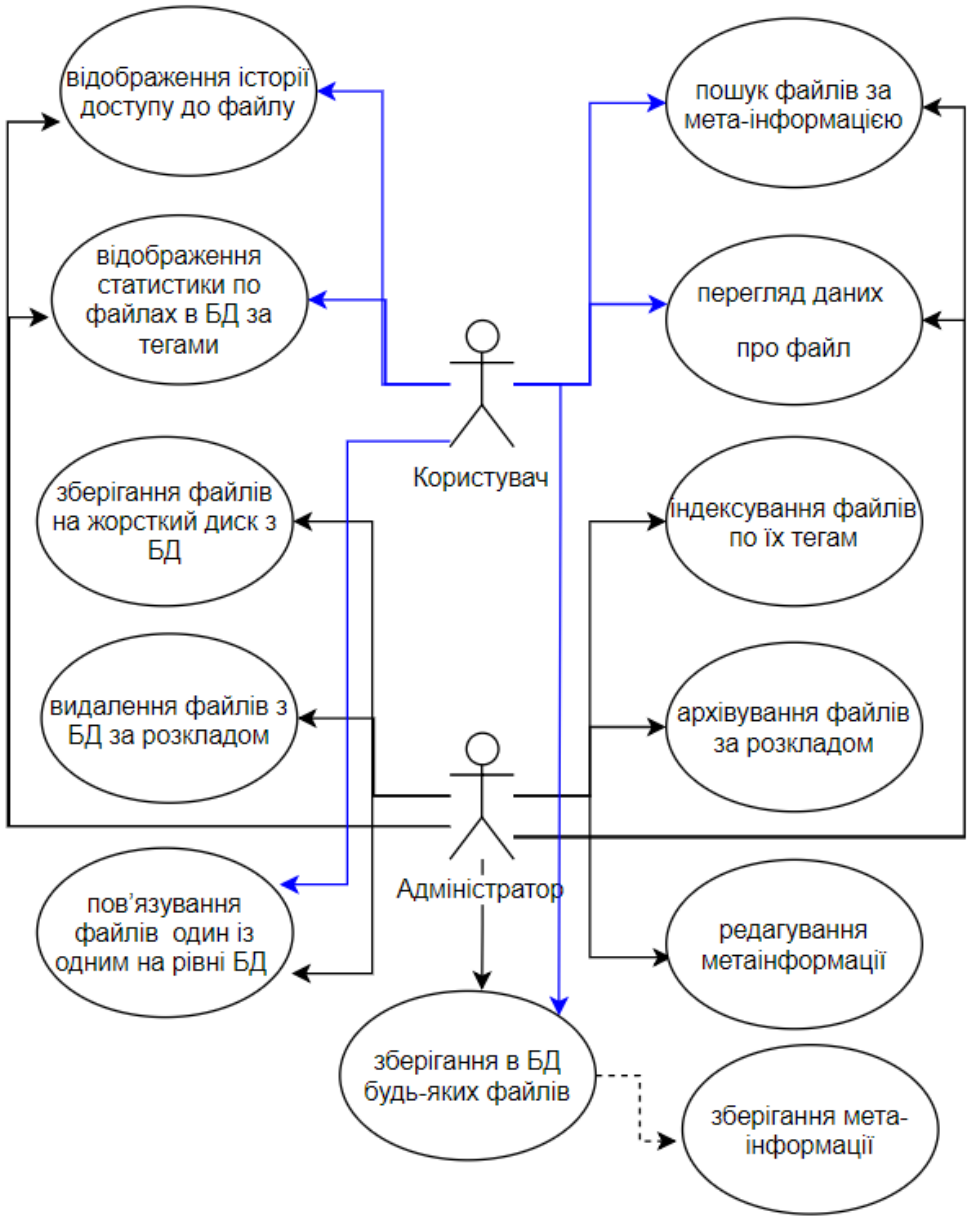
InterSystems iKnow – технологія, яка дозволяє розробникам створювати застосування для отримання інформації з неструктурованих даних . Технологія iKnow дозволяє знаходити потрібні концепти і зв'язки в неструктурованих даних.

Програмний продукт реалізований *мовою C#* засобами Visual Studio з використанням об'єктного доступу до даних, що зберігаються в СКБД.

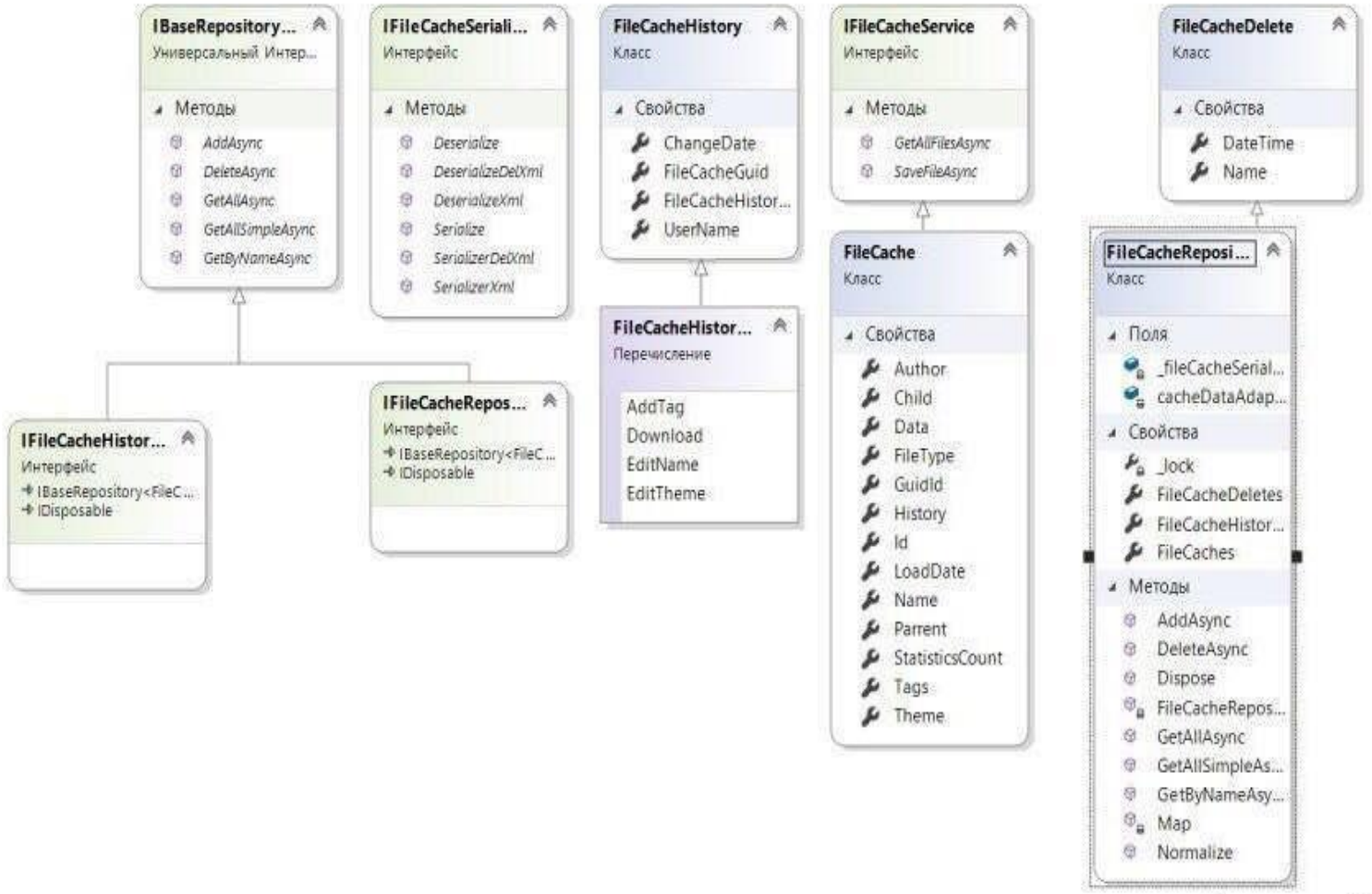
Архітектура ПЗ



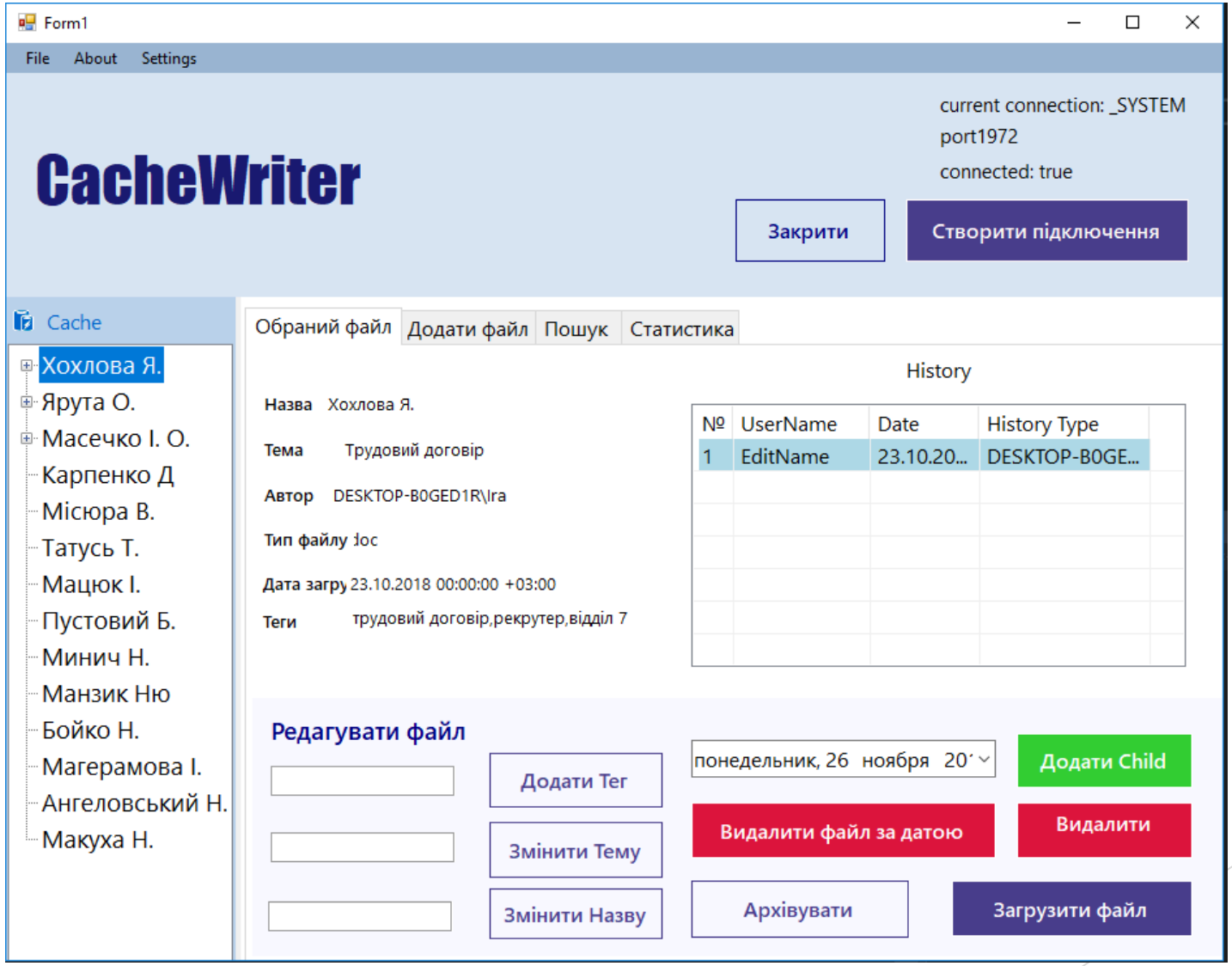
Діаграма прецедентів



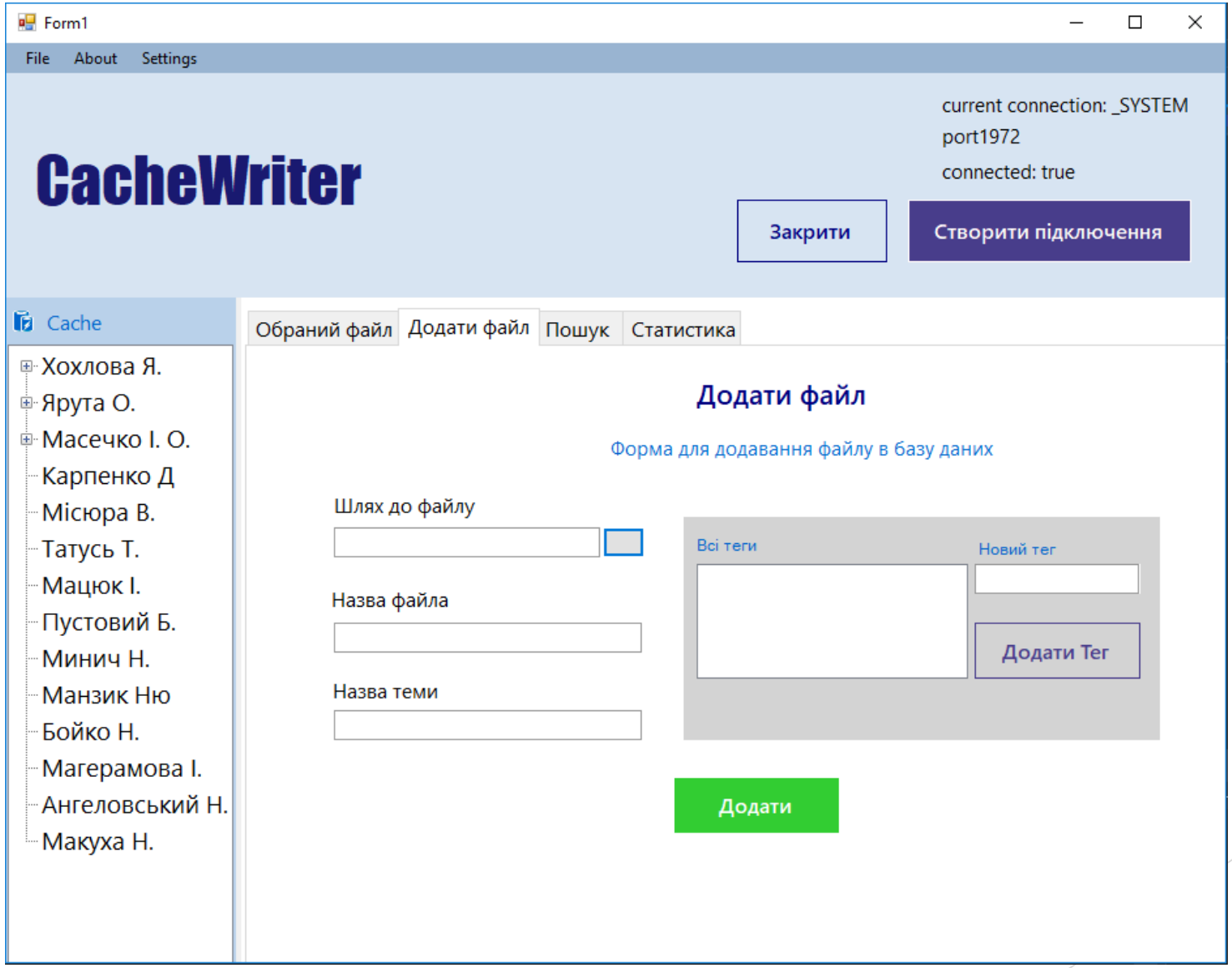
Структура програмної системи



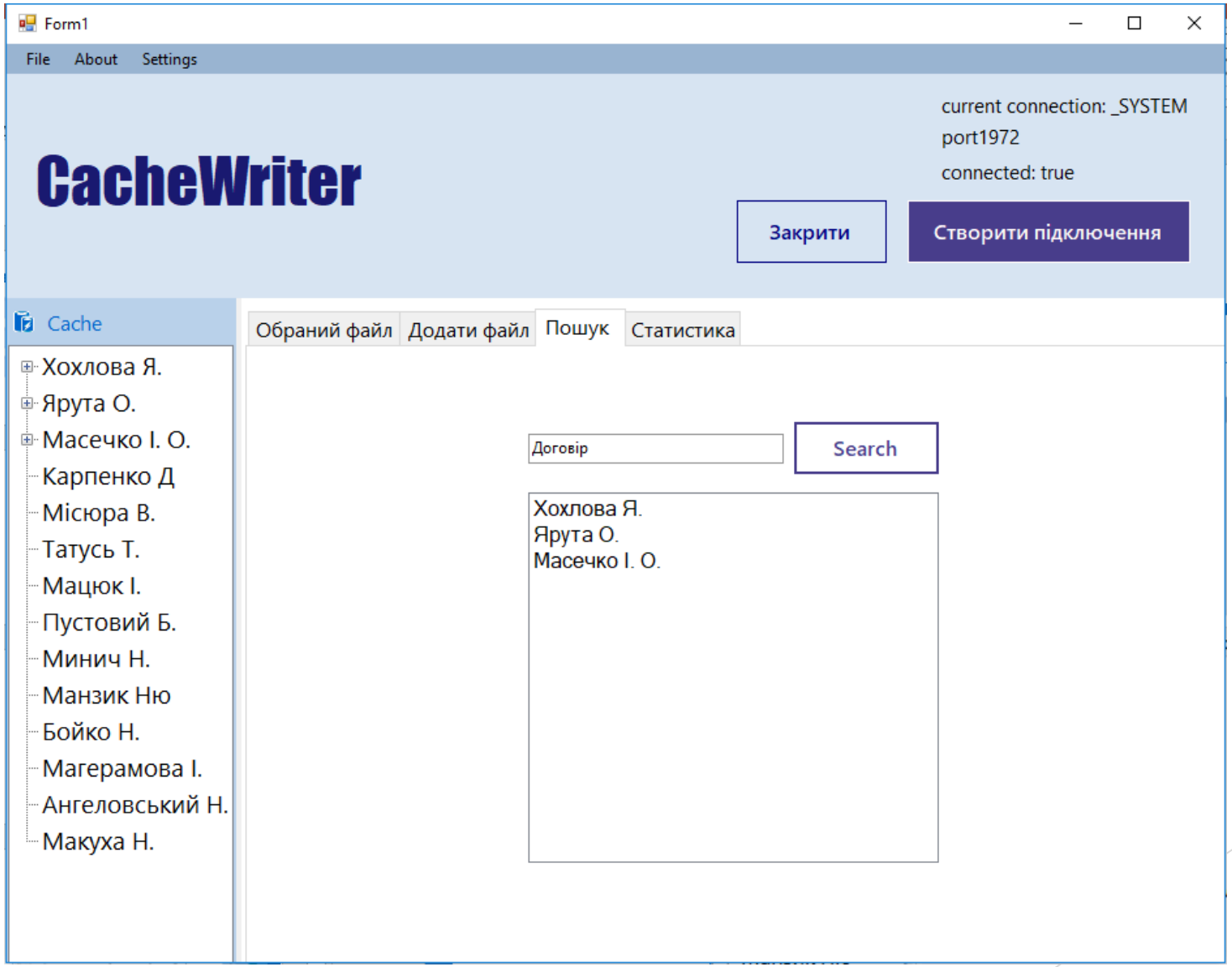
Приклад роботи програми



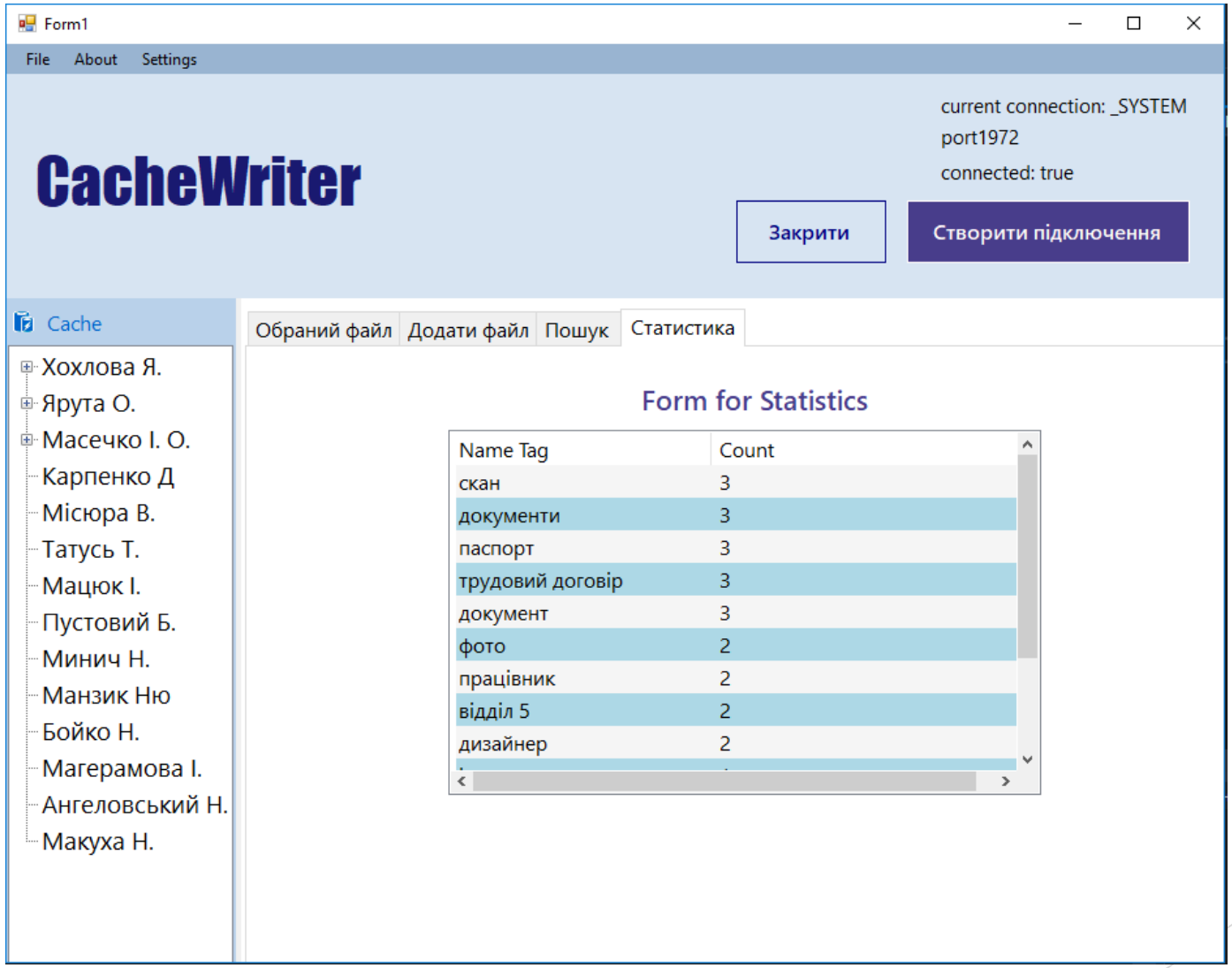
Приклад роботи програми



Приклад роботи програми



Приклад роботи програми



Висновки

- Під час виконання магістерської дисертації була проаналізована та вивчена проблема роботи з неструктурованими документами. Було встановлено, що для вирішення завдання зберігання та обробки неструктурованих документів, недостатньо існуючих на сьогодні програмних рішень.
- Було проведено огляд методів і засобів розробки програмної системи для вирішення даного завдання.
- Розроблено архітектуру програмного забезпечення для зберігання та обробки неструктурованих даних. Розроблено загальну методику роботи користувача з програмою, що забезпечує швидку адаптацію користувача до особливостей роботи застосунку.
- Було спроектовано та реалізовано програмну систему роботи з неструктурованими даними з використанням технологій InterSystems, яка надає можливість з легкістю маніпулювати великою кількістю документів та виконувати пошук по ним.
- Розроблений програмний продукт виконано відповідно до заданих вимог, з використанням патерну Repository та сучасних підходів до розробки програмних застосунків.
- Система впроваджена та використовується працівниками компанії ТОВ "МОБІМІЛЛ СОФТ".

Дякую за увагу