

# Розробка інтелектуальної системи інтернет-парсингу контенту веб-ресурсів

Розробив студент групи 1КН-14сп  
Кратасюк Віктор Васильович  
Керівник – Арсенюк Ігор Ростиславович

Вінниця 2015р

- **Об'єктом** дослідження є процес аналізу та отримання інформації з веб-ресурсів.
- **Предметом** дослідження є моделі та методи отримання контенту веб-ресурсів з використанням програмних засобів.
- **Метою** дослідження є покращення якості отримання інформації, що розміщується на веб-ресурсах шляхом розроблення інтелектуальної системи інтернет-парсингу контенту веб-ресурсів.

# Актуальність роботи

---

Розроблена програма може використовуватись:

- Для вилучення контенту із веб-ресурсів представлених в мережі Інтернет;
- Зручного перенесення інформації із застарілого сайту на оновлений ресурс;
- Як засіб збору інформації для наповнення веб-сайту.

# Аналіз методів вирішення задачі

---

В ході роботи розглядалися наступні методи:

- Метод, що базується на **LL** аналізаторі ;
- Метод, що базується на **RL** аналізаторі ;
- Метод рекурсивного спуску;
- Метод на базі регулярних виразів;
- Метод, що базується на генетичному алгоритмі.

# Удосконалення математичної моделі

---

Для усунення збіжності було використано:

- Лінійне масштабування полягає у перетворенні функції пристосованості  $F$  до форми  $F'$  через лінійну залежність виду:  $F' = AF + B$ ,
- Сигма-відсікання - метод масштабування, заснований на перетворенні функції пристосованості  $F$  до форми  $F'$  згідно з виразом:  $F' = F + (\bar{F} - C\sigma)$ ,
- Ступінне масштабування - являє собою метод масштабування, при якому функція пристосованості  $F$  перетвориться у форму  $F'$  згідно з виразом:  $F' = F^k$ ,

# Для надання функції пристосованості адаптивності було використано:

Припустимо, що ми маємо точку  $\bar{x}$ , для якої буде проведена адаптація.  
Нехай  $F_0(\mathbf{ch}_i)$  – поточне значення функції пристосованості.

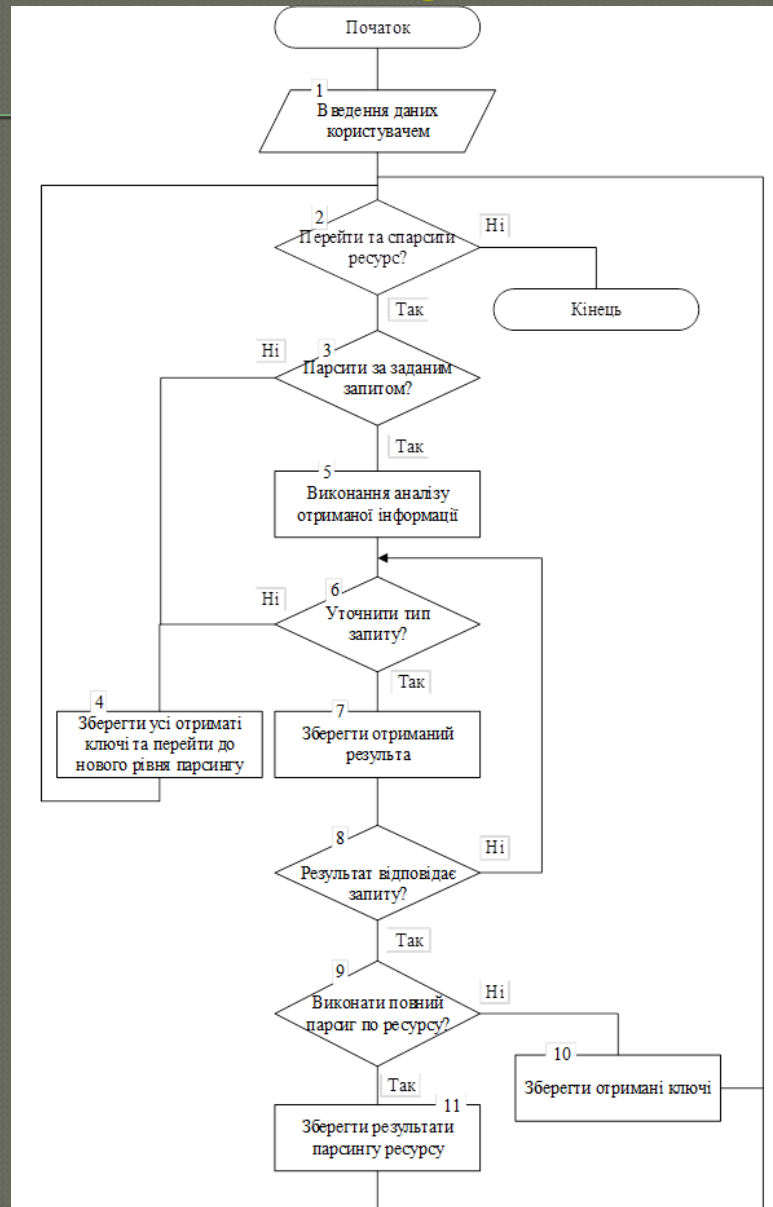
1) Якщо  $\bar{x}$  – локальне оптимальне рішення, тоді функція пристосованості буде модифікована згідно з функцією тунелювання:

$$F(x, \bar{x}) = F_0(x) \cdot \exp\left(\frac{1}{p + \frac{1}{r^2} \|x - \bar{x}\|^2}\right).$$

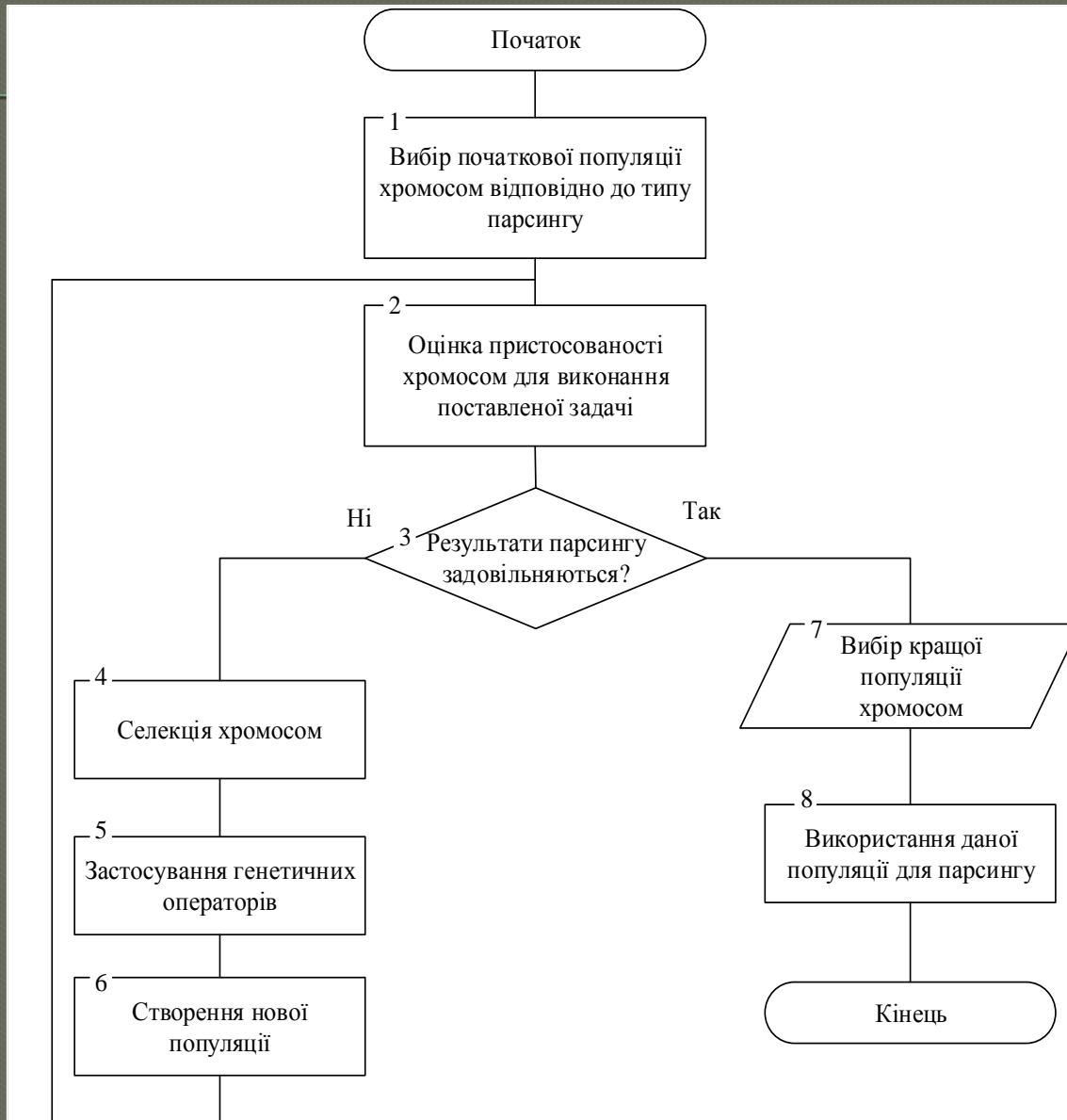
2) Якщо  $\bar{x}$  – глобальне оптимальне рішення, тоді функція пристосованості буде модифікована згідно з функцією горбоподібного тунелювання:

$$F_{gl}(x, \bar{x}) = F(x, \bar{x}) \cdot \exp\left(\frac{1}{p + \frac{1}{r^2} \|x - \bar{x}\|^2}\right) = \left(F_0(x) + \max\left\{0, 1 - \frac{1}{r^2} \|x - \bar{x}\|^2\right\}\right) \cdot \exp\left(\frac{1}{p + \frac{1}{r^2} \|x - \bar{x}\|^2}\right).$$

# Алгоритм інтелектуального модуля

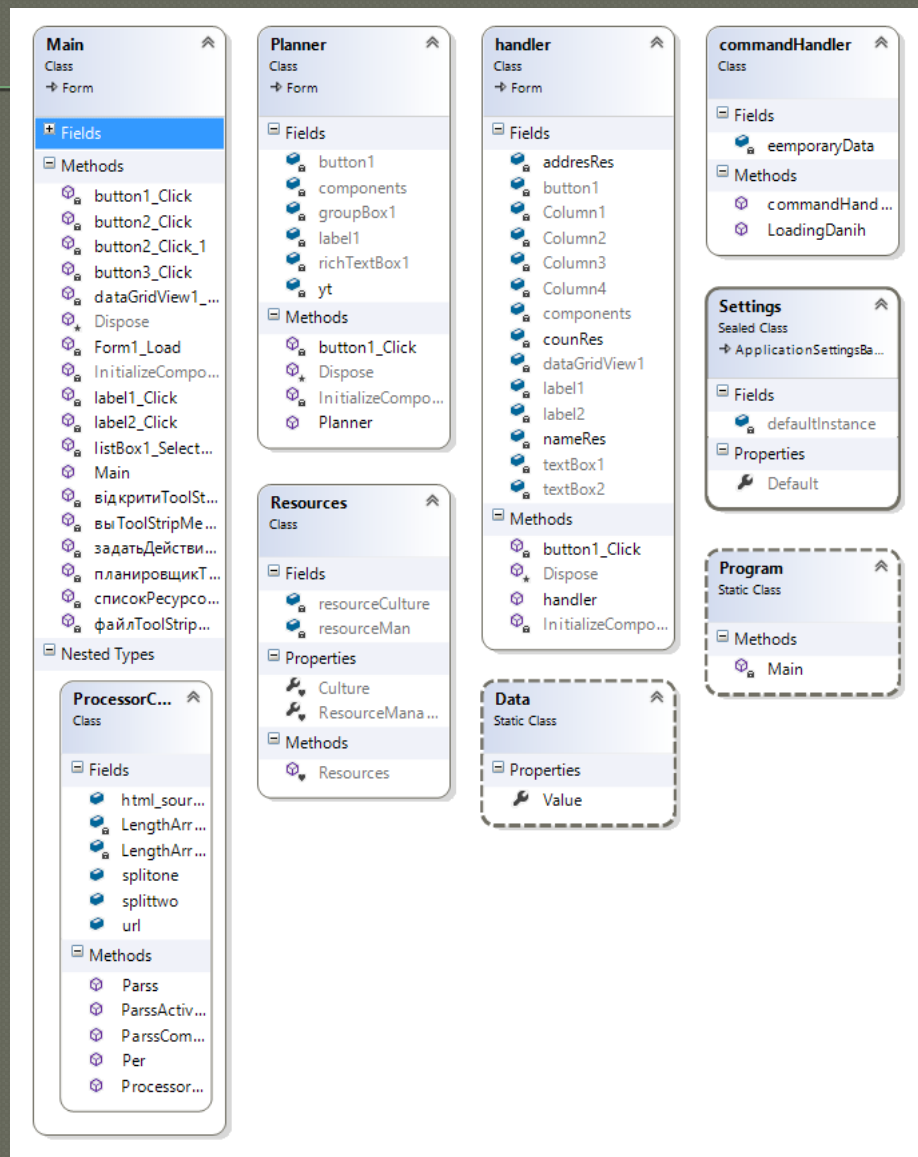


# Схема генетичного алгоритму

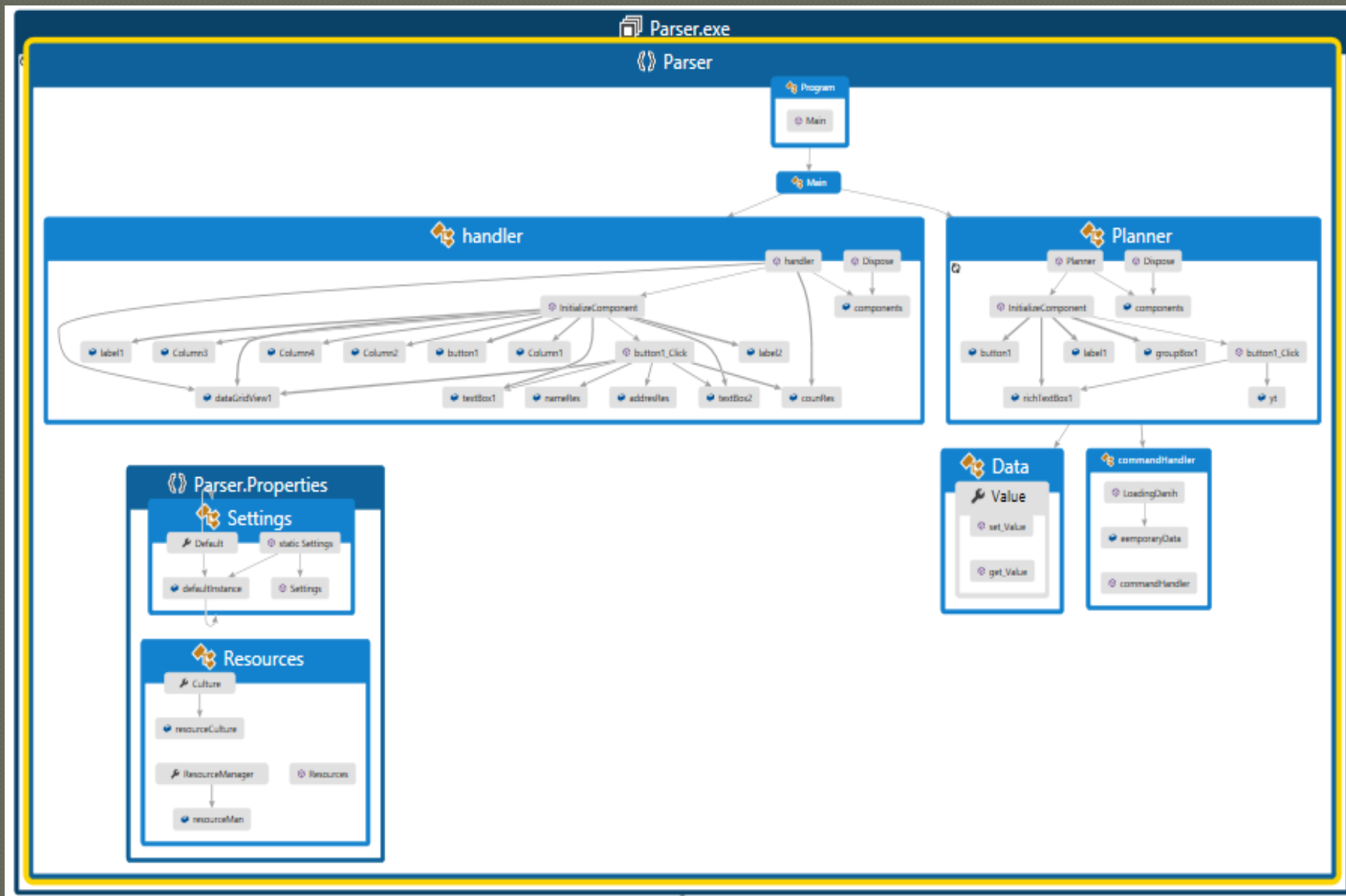




# UML діаграма

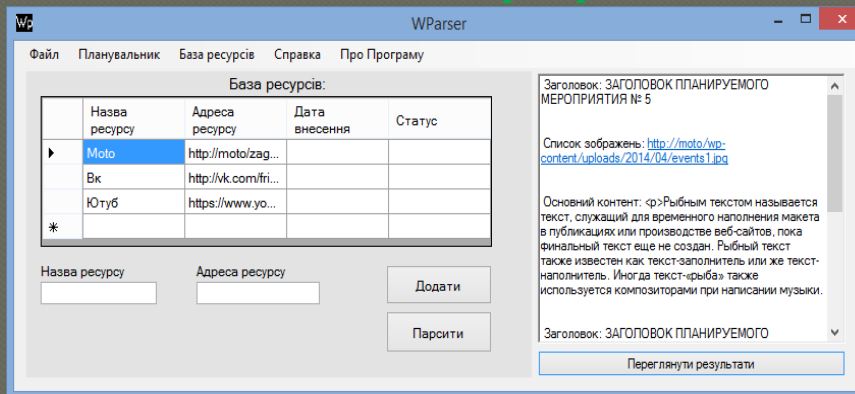


# Загальна схема взаємодії елементів

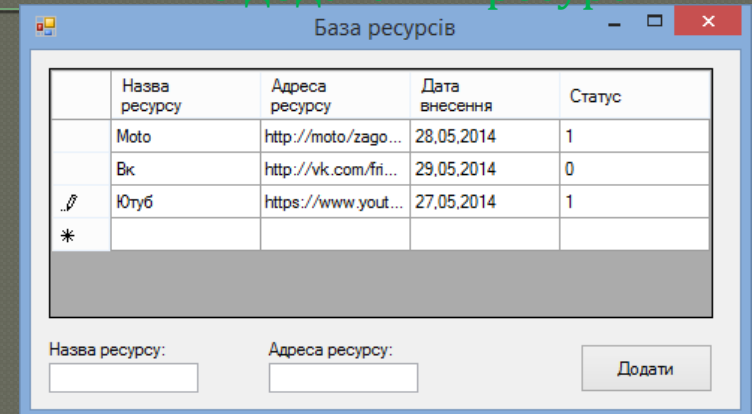


# Графічний інтерфейс програми

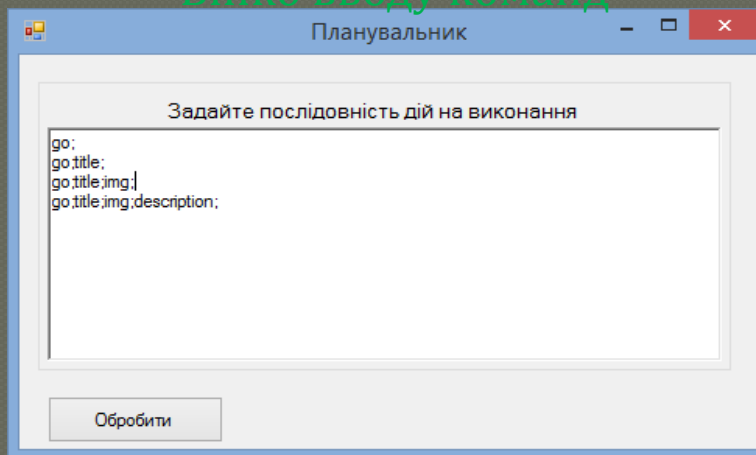
Головне вікно програми



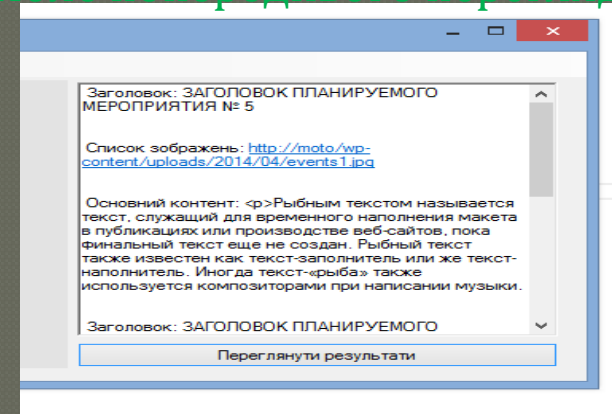
Вікно додавання ресурсів



Вікно вводу команд

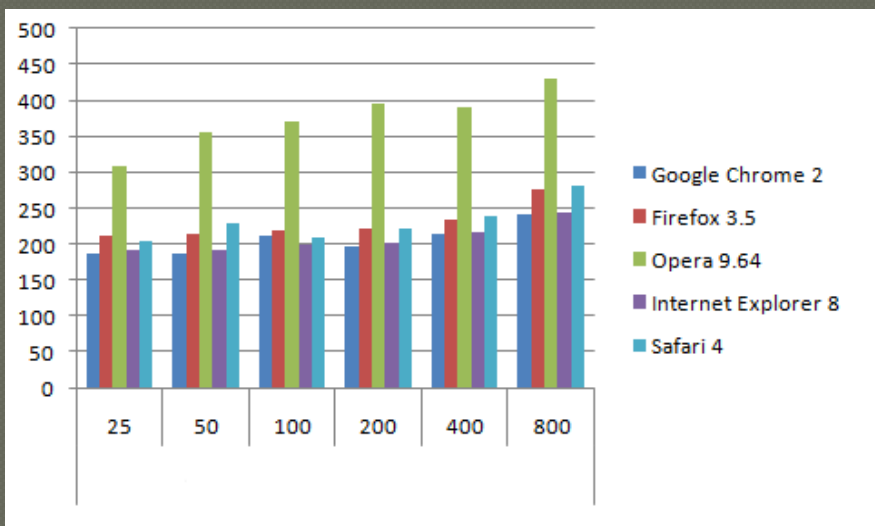


Поле попереднього перегляду

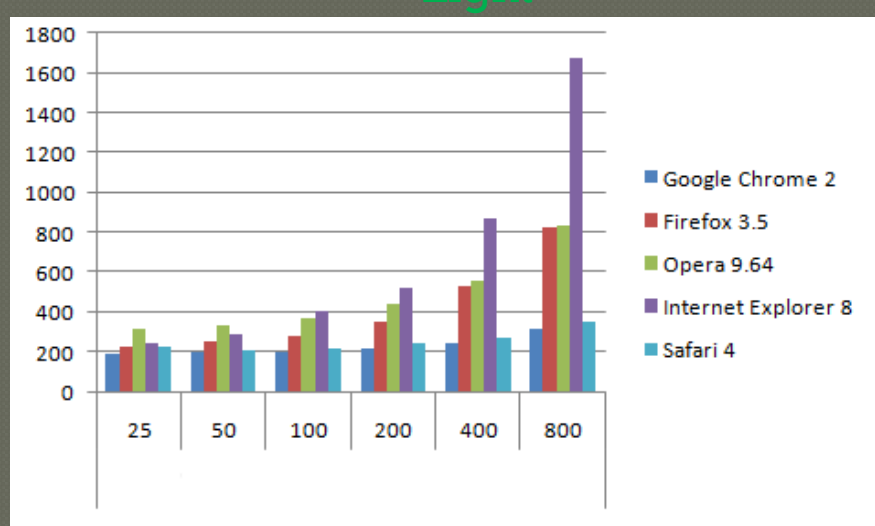


# Результати тестування

## Результати проектованої системи



## Результати аналога X-Parser Light



# Апробація роботи

Дана робота не одноразово доповідалася на науково-технічних конференціях професорсько-викладацького складу, співробітників та студентів університету за участю працівників науково-дослідних організацій та інженерно-технічних працівників підприємств м. Вінниці.

Має 4 публікації: *авторське право на твір* (програму), *матеріали міжнародної науково-практичної конференції „Інтернет – Освіта – Наука” та двоє тез доповідей вище вказаних конференцій.*



## Висновки:

Проведено дослідження та аналіз існуючих аналогів на ринку, та обґрунтування доцільності розробки інтелектуального модуля. Виконано обґрунтування вибору математичної моделі та її вдосконалення. Розроблено алгоритм інтелектуальної системи та обґрунтовано вибір генетичного алгоритму.

Проведені економічні розрахунки довели доцільність та окупність нової розробки. Робота містить розгляд детального планування та розробки об'єктно орієнтованої системи з використанням UML діаграм. Після розв'язання усіх поставлених задач та аналізу результатів тестування системи мета роботи була досягнута в повній мірі.

---

Дякую за увагу!