

ДОДАТОК А

Програмний код для морфологічних правил

Програмний код для правила з'єднання субстантивного словосполучення та кількісного словосполучення.

```

if (NumP.вид == «порядковий»)
{
    if (NumP.число == NounP.число)
    {
        if (NumP.число == «однина»)
        {
            if (NumP.рід == NounP.рід && NumP.відмінок ==
NounP.відмінок)
                return true;
            else
                return false;}
        if (NumP.число == «множина»)
        {
            if (NumP.відмінок == NounP.відмінок)
                return true;
            else
                return false;}}
    else
        return false;}
if (NumP.вид == «кількісний»)
{
    if (NumP == «один»)
    {
        if (NumP.число == NounP.число && NumP.рід ==
NounP.рід && NumP.відмінок == NounP.відмінок)
            return true;
        else
            return false;}
    else if ( NumP == «два» || NumP == «оба»)
    {
        if (NumP.рід == NounP.рід && NumP.відмінок ==
NounP.відмінок)
            return true;
        else
            return false;}
    }else{

```

```
if(NumP.відмінок == NounP.відмінок)
    return true;
else
    return false;}}
```

Програмний код для правила, яке містить займенникове словосполучення разом з субстантивним.

```
if(PropP.число == NounP.число)
{
    if(PropP.число == «однина»)
    {
        if(PropP.під == NounP.під && PropP.відмінок ==
NounP.відмінок)
            return true;
        else
            return false;
    }
    if(PropP.число == «множина»)
    {
        if(PropP.відмінок == NounP.відмінок)
            return true;
        else
            return false;
    }
}
else
    return false;
```

ДОДАТОК Б
Слайди презентації

1

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Атестаційна робота магістра

Дослідження методів синтаксичного аналізу для системи
автоматичної обробки текстової інформації

Науковий керівник:
к.т.н., доц., каф. ПІ
Валенда Н.А.

Виконав:
студент групи ПЗСм-18-1
Долуда В.П.

Рисунок Б.1 – Титульний слайд

ОСНОВНІ ВІДОМОСТІ ПРО ДОСЛІДЖЕННЯ

- ▶ Об'єктом дослідження є процес обробки текстової інформації.
- ▶ Метою роботи є дослідження проблем синтаксичного аналізу та розробка методу синтаксичного аналізу для мов з нечітким порядком слів у реченні.
- ▶ Наукова новизна роботи полягає у комбінованому підході до синтаксичного аналізу, який включає синтаксичний аналіз для груп слів з чітким порядком слів, та елементи семантичного аналізу для усього речення в цілому.

Рисунок Б.2 – Основні відомості про дослідження

3

Постановка задачі

- ▶ Формалізувати синтаксис словосполучень – побудувати модель, яка в формальному вигляді опише синтаксис даних мовних одиниць;
- ▶ Запропонувати механізм співставлення граматичних ознак головних та залежних слів;
- ▶ Запропонувати метод семантико-синтаксичного аналізу речень на основі словника валентності.

Рисунок Б.3 – Постановка задачі

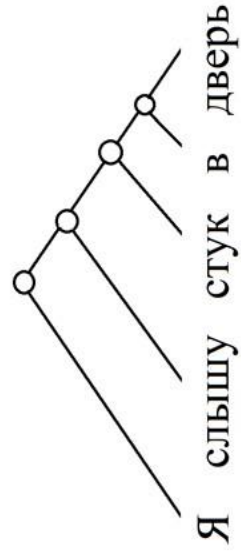


Рисунок Б.4 – Етапи лінгвістичного процесору

5

СИНТАКСИЧНИЙ РОЗБІР РЕЧЕННЯ

Дерево складових



Дерево зависимостей

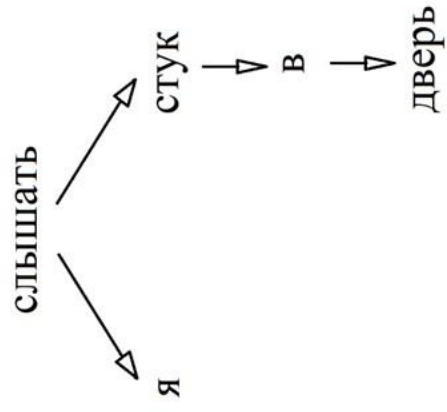


Рисунок Б.5 – Синтаксичний розбір речення



Рисунок Б.6 – Словники валентності

ВИДИ СЛОВОСПОЛУЧЕНЬ

Вид словосполучення	Підвид	Головне слово	Залежне слово	Тип зв'язка
Іменні	Субстантивні	Іменник	Іменник	Управління
			Прикметник	Узгодження
			Займенник	Узгодження
			Числівник	Узгодження
			Причастя	Узгодження
			Інфінітив	Примикання

Рисунок Б.7 – Види словосполучень

КОНТЕКСТНО-ВІЛЬНА ГРАМАТИКА

$$G = (V_t, V_n, P, S)$$

- $V_t =$
- ▶ verb (дієслово);
 - ▶ noun (іменник);
 - ▶ adj (прикметник);
 - ▶ adv (прислівник);
 - ▶ inf (інфінітив);
 - ▶ pron (займенник);
 - ▶ part (причастя);
 - ▶ numeral (числівник);
 - ▶ gerund (дієприслівник);
 - ▶ prepos (прийменник).

$$V_n =$$

- ▶ VerbP (дієслівне словосполучення);
- ▶ NounP (субстантивне словосполучення);
- ▶ PronP (займенникове словосполучення);
- ▶ NumP (кількісне словосполучення);
- ▶ AdjP (ад'єктивне словосполучення);
- ▶ AdvP (прислівникове словосполучення);
- ▶ PreP (прийменник);
- ▶ Phrase (словосполучення).

Рисунок Б.8 – Контекстно-вільна граматики

Правила грамматики

$P = \{$

- (1) $Phrase \rightarrow \langle VerbP \rangle | \langle NounP \rangle | \langle PronP \rangle | \langle NumP \rangle | \langle AdjP \rangle | \langle AdvP \rangle$,
 - (2) $VerbP \rightarrow \langle VerbP \rangle \langle NounP \rangle | \langle VerbP \rangle \langle AdjP \rangle | \langle VerbP \rangle \langle PronP \rangle | \langle VerbP \rangle \langle AdvP \rangle$,
 - (3) $VerbP \rightarrow \langle VerbP \rangle gerund | \langle VerbP \rangle inf | \langle Prep \rangle verb$,
 - (4) $NounP \rightarrow \langle AdjP \rangle \langle NounP \rangle | \langle NounP \rangle \langle NounP \rangle | part \langle NounP \rangle | \langle NumP \rangle \langle NounP \rangle | \langle PronP \rangle \langle NounP \rangle$,
 - (5) $NounP \rightarrow \langle Prep \rangle noun$,
 - (6) $PronP \rightarrow \langle PronP \rangle \langle PronP \rangle | \langle PronP \rangle \langle AdjP \rangle | \langle Prep \rangle pron$,
 - (7) $NumP \rightarrow \langle NumP \rangle \langle NounP \rangle | \langle NumP \rangle \langle AdjP \rangle | \langle NumP \rangle part | \langle Prep \rangle numeral$,
 - (8) $AdjP \rightarrow \langle AdjP \rangle \langle NounP \rangle | \langle AdjP \rangle inf | \langle AdjP \rangle \langle AdvP \rangle | \langle Prep \rangle adj$,
 - (9) $AdvP \rightarrow \langle AdvP \rangle \langle NounP \rangle | \langle AdvP \rangle \langle AdjP \rangle | \langle AdvP \rangle \langle AdvP \rangle | \langle Prep \rangle adv$,
 - (10) $Prep \rightarrow prepos | \epsilon$
- $\};$

Рисунок Б.9 – Правила грамматики

Морфологічні Правила

NounP → < *AdjP* > < *NounP* >

10

```
if(AdjP.число == NounP.число)
{
  if(AdjP.число == «однина»)
  {
    if(AdjP.pіa == NounP.pіa &&
      AdjP.вiдмінок == NounP.вiдмінок)
      return true;
    else
      return false;
  }
  if(AdjP.число == «множина»)
  {
    if(AdjP.вiдмінок == NounP.вiдмінок)
      return true;
    else
      return false;
  }
}
else
  return false;
```

Рисунок Б.10 – Морфологічні правила

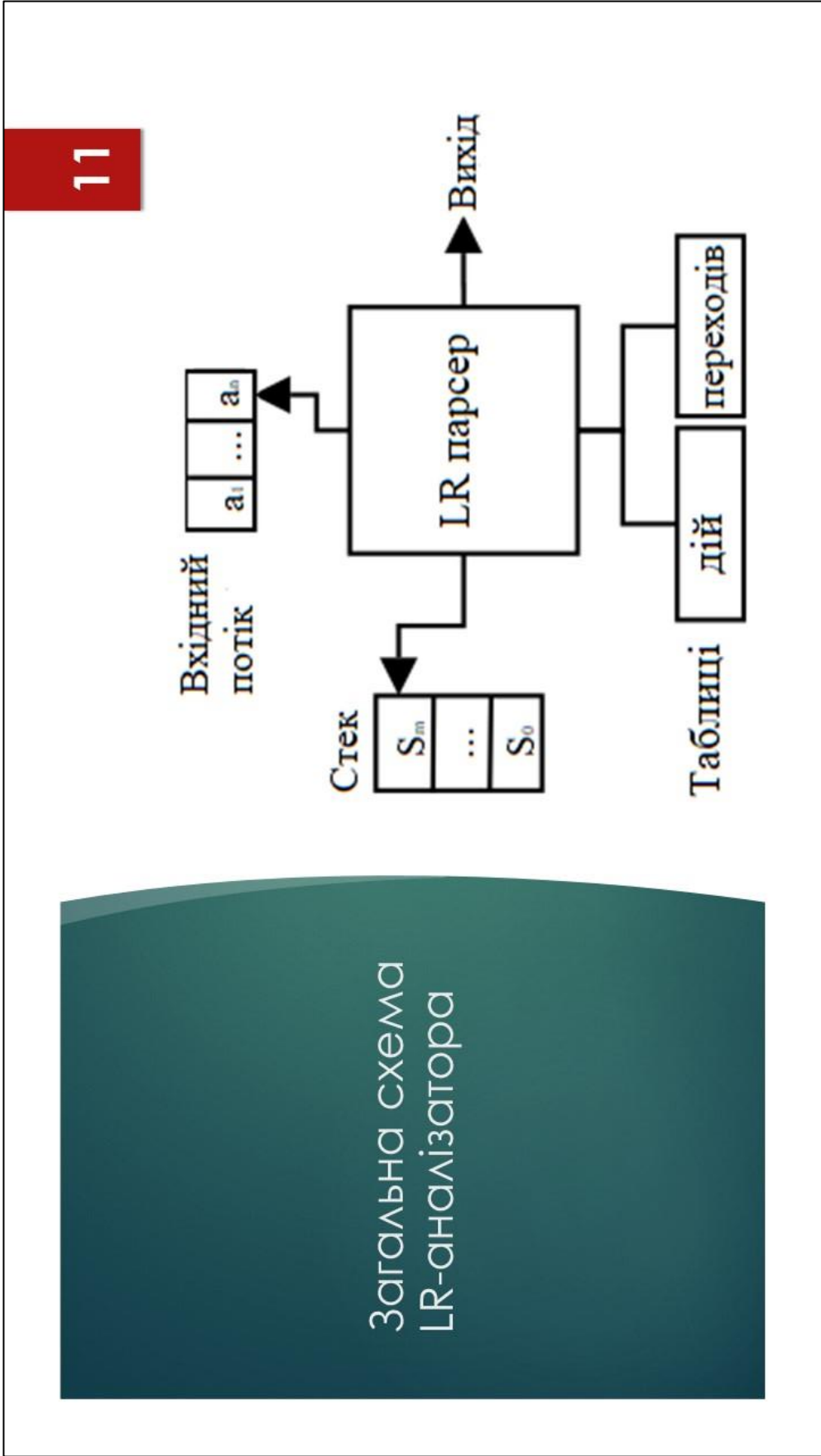


Рисунок Б.11 – Загальна схема LR-аналізатора

Алгоритм семантико-синтаксичного аналізатору

- ▲ Використовувати словник валентності.
- ▲ Пошук дієслівного словосполучення.
- ▲ Віднесення інших словосполучень до актантів дієслова.

Рисунок Б.12 – Алгоритм семантико-синтаксичного аналізатору

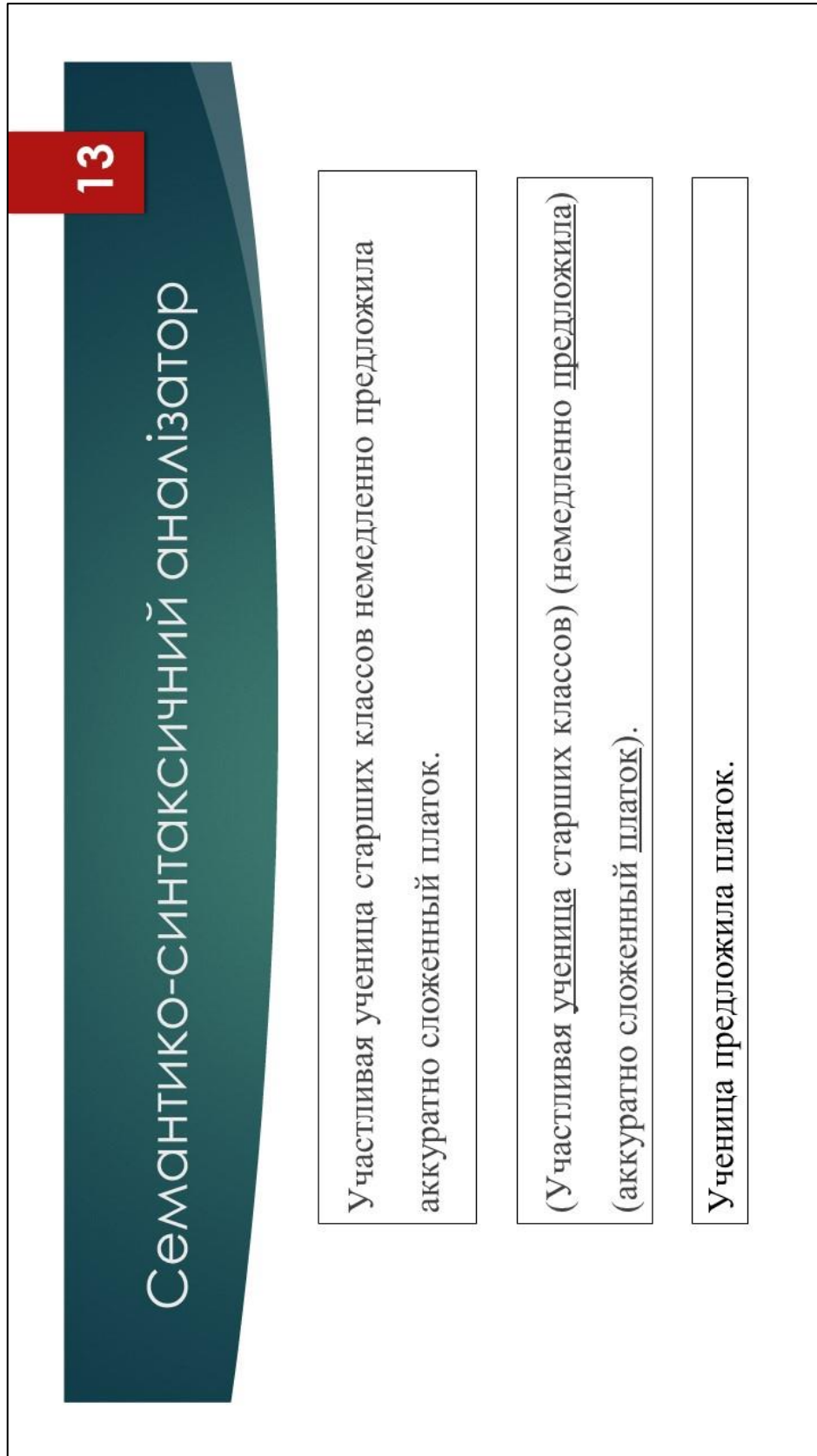


Рисунок Б.13 – Семантико-синтаксичний аналізатор

14

МОЖЛИВОСТІ ВИКОРИСТАННЯ

- ▶ Напівчорнове створення корпусів тексту.
- ▶ Пошук помилок при узгодженні частин мови.

Рисунок Б.14 – Можливості використання

15

Висновки

- ▲ Проаналізовано предметну галузь.
- ▲ Створено контекстно-вільну граматику для пошуку словосполучень.
- ▲ Запропоновано метод семантико-синтаксичного аналізу речень на основі словника валентності.
- ▲ Створено парсер для пошуку словосполучень.

Рисунок Б.15 – Висновки



Рисунок Б.16 – Дякую за увагу

ДОДАТОК В
Відгук керівника роботи

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ
Факультет комп'ютерних наук

ВІДГУК

на атестаційну роботу магістра
Долуди Валерії Павлівни, ПЗСм-18-1

спеціальність: 121-«Інженерія програмного забезпечення»
освітньо-професійна програма: «Програмне забезпечення систем»
Тема атестаційної роботи: «Дослідження методів синтаксичного аналізу для системи автоматичної обробки текстової інформації».


В ході проведення дослідження студентом було запропоновано новий підхід до вирішення проблеми синтаксичного аналізу речення. Наукова новизна роботи полягає у комбінованому підході до синтаксичного аналізу, який включає синтаксичний аналіз для груп слів з чітким порядком слів, та елементи семантичного аналізу для усього речення в цілому.

Також в ході виконання атестаційної роботи студент продемонстрував вміння роботи з науково-технічною літературою та відкритими джерелами в мережі Інтернет. Студент продемонстрував вміння шукати та аналізувати накопичену інформацію про синтаксичний аналіз текстової інформації та використовувати її в рамках дослідження.

За період виконання атестаційної роботи Долуда В.П. продемонструвала самостійність та високий рівень відповідальності. Усі етапи роботи були виконані вчасно та з використанням сучасних засобів дослідження та розробки.

Магістрант гр. ПЗСм-18-1 Долуда В.П. готовий до самостійної інженерної діяльності. Атестаційну роботу можна подати до захисту в ЕК за спеціальністю 121-«Інженерія програмного забезпечення», освітньо-професійною програмою «Програмне забезпечення систем».

« 17 » грудня 2019 р.


Керівник атестаційної роботи магістра
к.т.н. доц. Валенда Н.А.

ДОДАТОК Г

Рецензії

Рецензія

на атестаційну роботу магістра

студента групи ПЗСм-18-1: Долуди Валерії Павлівни

спеціальність: 121-«Інженерія програмного забезпечення»

освітньо-професійна програма: «Програмне забезпечення систем»

«Дослідження методів синтаксичного аналізу для системи автоматичної обробки
текстової інформації»

(Тема атестаційної роботи)

Структура атестаційної роботи: пояснювальна записка 77 сторінок.

У запропонованій для рецензування атестаційній роботі магістра проведено дослідження методів синтаксичного аналізу для системи автоматичної обробки текстової інформації.

Обсяг пояснювальної записки та її розділів відповідає вимогам. Матеріал поданий послідовно та структуровано. У межах роботи проаналізовані технічна література та дослідження зарубіжних вчених. Цитування використаних джерел адекватне тексту записки. Пояснювальна записка відповідає ДСТУ 3008:2015.

В ході виконання роботи було виявлено труднощі при створенні граматики для цілого речення. Тому було створено контекстно-вільну граматику для пошуку словосполучень російської мови.

Результати роботи можуть бути використані як частина семантико-синтаксичного аналізу. Також отриманий парсер можна використовувати для створення напівчорнових корпусів тексту.

До недоліків роботи треба відмітити роботу отриманої граматики лише на простих реченнях, в яких відсутні розділові знаки. Також через це граMATика не обробляє однорідні члени речення.

Атестаційна робота магістранта групи ПЗСм-18-1 Долуди В. П. відповідає вимогам до атестаційних робіт і заслуговує оцінки «добре – 80». Атестаційну роботу можна представити для захисту в ЕК за спеціальністю 121-«Інженерія програмного забезпечення», освітньо-професійною програмою «Програмне забезпечення систем».

Рецензент

к.т.н. доц. каф програмної інженерії



Назаров О.С.

Рецензія

на атестаційну роботу магістра

студента групи ПЗСм-18-1: Долуди Валерії Павлівни

спеціальність: 121-«Інженерія програмного забезпечення»

освітньо-професійна програма: «Програмне забезпечення систем»

«Дослідження методів синтаксичного аналізу для системи автоматичної обробки
текстової інформації»

(Тема атестаційної роботи)

Структура атестаційної роботи: пояснювальна записка 77 сторінки.

У запропонованій для рецензування атестаційній роботі магістра проведено дослідження методів синтаксичного аналізу для системи автоматичної обробки текстової інформації.

Пояснювальна записка повністю відповідає вимогам ДСТУ 3008:2015. Обсяг та структура подання матеріалу відповідає нормам. У межах роботи використана та проаналізована технічна література. Автор демонструє високий рівень обізнаності в роботах з дослідження синтаксичного аналізу.

В ході роботи магістрант робить детальний аналіз існуючих підходів до синтаксичного аналізу та існуючих синтаксичних парсерів. У ході аналізу магістрант робить висновки щодо неефективності написання граматик для розбору речень цілком та пропонує підхід, який об'єднує в собі написання граматики словосполучення та використання словників валентності.

В результаті роботи було створено контекстно-вільну граматику словосполучень та реалізовано парсер на її основі. Робота виконана на високому науковому рівні. Надалі результати роботи можуть бути використані як частина семантико-синтаксичного аналізу.

Серед недоліків роботи відсутність критеріїв для оцінки роботи отриманого парсера. Для перевірки отриманого результату необхідно порівняння результатів роботи парсера з аналогами.

Атестаційна робота магістранта групи ПЗСм-18-1 Долуди В.П. відповідає вимогам до атестаційних робіт і заслуговує оцінки «добре – 75». Атестаційну роботу можна представити для захисту в ЕК за спеціальністю 121-«Інженерія програмного забезпечення», освітньо-професійною програмою «Програмне забезпечення систем».

Рецензент



доцент каф. СТ Реберак А.М.