

Міністерство освіти і науки України  
Національний університет «Львівська політехніка»

Марків Оксана Олександрівна



УДК 81'322.2:[004.774+004.773.7+004.624]

**ЛІНГВІСТИЧНЕ ЗАБЕЗПЕЧЕННЯ  
ФОРМУВАННЯ ТУРИСТИЧНОЇ ДОКУМЕНТАЦІЇ  
НА ОСНОВІ ВІДКРИТИХ ВЕБ-РЕСУРСІВ**

10.02.21 – структурна, прикладна і математична лінгвістика

**Автореферат**  
дисертації на здобуття наукового ступеня  
кандидата технічних наук

Львів – 2017

Дисертацією є рукопис.

Робота виконана у Національному університеті «Львівська політехніка» Міністерства освіти і науки України, м. Львів.

**Науковий керівник:** доктор технічних наук, професор  
**Жежнич Павло Іванович**,  
Національний університет «Львівська політехніка»,  
професор кафедри соціальних комунікацій  
та інформаційної діяльності

**Офіційні опоненти:** доктор технічних наук, професор  
**Виклюк Ярослав Ігорович**,  
проректор з наукової роботи та міжнародних  
відносин Буковинського університету, м. Чернівці

кандидат технічних наук  
**Нога Роман Юрійович**,  
провідний інженер ТзОВ «Levi Nine Ukraine»

Захист відбудеться 06 квітня 2017 року, о 16 год. на засіданні спеціалізованої вченої ради Д 35.052.05 у Національному університеті «Львівська політехніка» (79013, м. Львів, вул. С. Бандери, 12, ауд. 226 гол. корпусу).

З дисертацією можна ознайомитися у бібліотеці Національного університету «Львівська політехніка» (79013, м. Львів, вул. Професорська, 1).

Автореферат розісланий 04 березня 2017 року.

Учений секретар спеціалізованої  
вченої ради, д.т.н., професор



Р.А.Бунь

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність теми.** На сьогодні туризм є одним із перспективних напрямів як внутрішньо-, так і зовнішньоекономічної діяльності країни та її регіонів. Аналіз туристичної діяльності показує, що за останні 25 років кількість міжнародних туристів у світі зросла майже втричі. В Україні динаміка росту міжнародних туристів до 2014 була ще стрімкішою.

Основним джерелом інформації про туристичну діяльність є звичайно Всесвітня павутина WWW. Експоненціальний ріст обсягів інформації, що є доступною у WWW, призводить до того, що з кожним роком туристична інформація потребує все більших затрат часу та людських ресурсів на її опрацювання.

Дослідження щодо використання новітніх інформаційних технологій у діяльності підприємств туристичної сфери і туризмі в цілому здійснювали С.Арімов, В.Балута, П.Бурцев та інші. Моделюванню об'єктів туристичної галузі за допомогою інформаційних технологій присвячено роботи Я.І.Виклюка. Однак ці дослідження не розглядають на науковому рівні питання використання інформаційного наповнення веб-сайтів. Це призводить до того, що туристичні компанії не належним чином використовують інформаційні ресурси WWW, які формуються як надавачами туристичних чи супутніх послуг, так і їх споживачами.

Основна проблема масового використання інформаційного наповнення веб-сайтів для підтримки туристичної діяльності пов'язана з великою неоднорідністю структури веб-сторінок і слабкою структурованістю їхніх фрагментів. Тому акцент необхідно робити не на механістичному опрацюванні інформаційного наповнення веб-сторінок, а на лінгвістичних особливостях цього інформаційного наповнення з урахуванням таких чинників: інформація про туристичні послуги є розпорошеною на різних веб-сайтах; споживачі туристичних послуг надають окрему необхідну туристичну інформацію навіть краще, ніж надавачі туристичних послуг; туристична інформація вимагає постійної підтримки щодо повноти та достовірності.

Фактично туристична діяльність потребує якісного інформаційного забезпечення, яке формується як певний інформаційний продукт, який повинен задовольняти потреби споживача. Цей інформаційний продукт становить туристичну документацію, що містить інформацію з різних веб-ресурсів, згруповану за розділами.

Таким чином, сучасні туристичні підприємства потребують якісної документації для інформаційної підтримки надання широкого спектру туристичних послуг, яка б формувалася шляхом лінгвістичного аналізу інформаційного наповнення великої кількості веб-сайтів з відкритим доступом у WWW. Це підтверджує актуальність наукового завдання розроблення методів і засобів формування туристичної документації шляхом комп'ютерно-лінгвістичного аналізу інформаційного наповнення відкритих веб-ресурсів.

**Зв'язок роботи з науковими програмами, планами, темами.** Дисертаційні дослідження виконані у межах пріоритетного наукового напрямку розвитку науки і техніки "Інформаційні та комунікаційні технології", визначеного Законом України "Про пріоритетні напрями розвитку науки і техніки" від 05.12.2012 р. Тема дисертаційної роботи відповідає науковому напрямку кафедри інформаційних систем

та мереж Національного університету “Львівська політехніка”. Робота виконана в межах науково-дослідних робіт “Методи та засоби консолідації баз даних в інформаційних системах електронного урядування”, 2010-2012 рр., № держреєстрації 0110U005022 та “Лінгвістичне забезпечення консолідації відкритих інформаційних ресурсів”, 2013 р., № держреєстрації 0113U005274 (автором розроблено нові моделі та методи збору та верифікації туристичної інформації на основі комп'ютерно-лінгвістичного аналізу інформаційного наповнення неоднорідних відкритих веб-ресурсів).

**Мета і задачі дослідження.** Метою дослідження є розроблення методів і засобів формування туристичної документації за допомогою комп'ютерно-лінгвістичного аналізу інформаційного наповнення відкритих веб-ресурсів.

Для досягнення поставленої мети необхідно розв'язати такі задачі:

- проаналізувати особливості формування туристичної документації на основі методів структурування інформаційного наповнення відкритих веб-ресурсів;
- розробити інформаційну модель туристичної документації та інформаційного джерела з урахуванням лінгвістичних особливостей їх елементів;
- розробити метод лінгвістичного співставлення інформаційного наповнення відкритих веб-ресурсів з елементами туристичної документації;
- розробити процедури формування та виявлення недоліків туристичної документації на основі лінгвістичного співставлення інформаційного наповнення відкритих веб-ресурсів з елементами туристичної документації;
- розробити метод оцінювання якості туристичної документації;
- провести експериментальне дослідження та здійснити апробацію отриманих результатів.

*Об'єктом дослідження* є процес формування туристичної документації.

*Предметом дослідження* є методи і засоби формування туристичної документації за допомогою комп'ютерно-лінгвістичного аналізу та опрацювання інформаційного наповнення, що міститься у відкритих веб-ресурсах.

**Методи дослідження.** Дослідження, виконані під час роботи над дисертацією, ґрунтуються на теоретико-множинних підходах, апараті теорії реляційних баз даних та теорії графів для вирішення завдань інформаційного моделювання туристичної документації та відкритих веб-ресурсів; методах прикладної та математичної лінгвістики – для виявлення вербальних характеристик інформаційного наповнення відкритих веб-ресурсів; апараті нечітких множин – для побудови методів співставлення інформаційного наповнення відкритих веб-ресурсів з структурними елементами туристичної документації за допомогою лінгвістичних індикативних ознак; методах інформаційного пошуку в Інтернет – для процедури вибору відкритих веб-ресурсів для формування туристичної документації; методах інтелектуального аналізу текстів та методах побудови штучних мов маркування текстів – для розроблення процедур лінгвістичного співставлення інформаційного наповнення відкритих веб-ресурсів з елементами туристичної документації; теорії множин, частотному аналізі, методах визначення подібності текстів – для

розроблення процедури виявлення та усунення недоліків у туристичній документації; метриках оцінювання якості програмних продуктів за стандартом ISO/IEC-25010 – для розроблення методів оцінювання якості туристичної документації; діаграмних засобах структурного моделювання – для проектування прототипу системи автоматизованого формування туристичної документації.

**Наукова новизна одержаних результатів** полягає науковому обґрунтуванні та розробленні методів і засобів комп'ютерно-лінгвістичного аналізу інформаційного наповнення відкритих веб-ресурсів для формування туристичної документації. При цьому отримано такі наукові результати:

- уперше побудовано формальну модель туристичної документації на основі множин туристичних послуг, відгуків споживачів туристичних послуг та класифікаторів з урахуванням відповідних індикативних ознак їх елементів, що стало основою для побудови процедур комп'ютерно-лінгвістичного формування туристичної документації і показників її якості туристичної документації;
- удосконалено формальну модель інформаційного джерела для формування туристичної документації шляхом виокремлення блоків і атомарних ситуацій туристичної інформації на веб-сторінках, що стало основою для побудови процедури вибору відкритих веб-ресурсів, процедур лінгвістичного співставлення вмісту відкритих веб-ресурсів з елементами туристичної документації;
- набув подальшого розвитку метод лінгвістичного виявлення інформації у веб-ресурсах шляхом співставлення атомарних ситуацій з елементами туристичної документації на основі модифікованих індикативних ознак, що дозволило побудувати відповідні процедури лінгвістичного співставлення вмісту відкритих веб-ресурсів з елементами туристичної документації;
- уперше розроблено метод оцінювання якості туристичної документації на основі функціональних показників якості, показників якості придатності до використання та показників якості супроводжуваності стандарту ISO/IEC-25010, що забезпечило практичне оцінювання результатів виконання процедур формування та перевірки туристичної документації.

**Практичне значення одержаних результатів** дисертаційної роботи зумовлено тим, що вони дають змогу автоматизовано формувати туристичну документацію, проводити її перевірку та оцінювати її якість із врахуванням потреб споживача. Зокрема, практично цінними є такі результати:

- розроблено процедури комп'ютерно-лінгвістичного формування туристичної документації на основі співставлення вмісту відкритих веб-ресурсів з елементами за допомогою індикативних ознак, а також процедури виявлення прогалин та некоректної інформації у туристичній документації на основі аналізу її структури та вмісту, які дозволяють покращити показники якості туристичної документації і є необхідними для побудови системи автоматизованого формування туристичної документації;

- розроблено показники якості туристичної документації на основі функціональних показників якості, показників якості придатності до використання та показників якості супроводжуваності стандарту ISO/IEC-25010, що забезпечило практичне оцінювання результатів виконання процедур формування та перевірки туристичної документації;
- розроблено архітектуру та функціональну структуру системи комп'ютерно-лінгвістичного формування туристичної документації на основі процедур формування та показників якості туристичної документації, що дає можливість на практиці реалізувати і впровадити цю інформаційну систему.

Практичне значення дисертаційної роботи підтверджується позитивним ефектом від застосування розроблених методів і засобів комп'ютерно-лінгвістичного аналізу інформаційного наповнення відкритих веб-ресурсів для формування туристичної документації. Результати дисертаційного дослідження впроваджено у туристичних компаніях “Королівські подорожі” та “Тур де Львів”, а також у Національному університеті “Львівська політехніка”. Використання результатів дисертаційної роботи підтверджено актами про впровадження.

Результати дисертаційної роботи використовують у навчальному процесі в Національному університеті «Львівська політехніка» у лекційних курсах дисциплін «Методи та засоби інженерії даних та знань», «Методи та засоби інтеграції даних» під час підготовки студентів за спеціальністю 122 «Комп'ютерні науки та інформаційні технології», що підтверджено відповідними актами.

**Особистий внесок здобувача.** Усі наукові результати дисертаційної роботи отримані автором самостійно. У друкованих працях, опублікованих у співавторстві, автору належать: [1, 11] – метод формування та моніторингу якості туристичної документації, [3, 8, 14] – підхід до інформаційного моделювання за допомогою редукції графів, [5, 17, 18] – метод автоматизованого формування туристичної документації на основі інтелектуального аналізу текстів, [6, 16] – компоненти якості туристичної документації, [7] – аналіз процесу усунення прогалин у документації, [9] – лінгвістичний аналіз структурних елементів туристичної документації та метод формування туристичної документації з урахуванням прогалин, [10] – аналіз особливостей автоматизації опрацювання електронної документації у сфері надання туристичних послуг, [15] – метод формування туристичної документації на основі відкритих веб-ресурсів, [19] – технологія формування туристичної документації на основі слабоструктурованих текстів, [20] – підхід семантичного маркування слабоструктурованих даних [22] – аналіз особливостей генерування запитів у веб-системах, [21, 23] – множина соціо-культурних виразів, [24] – класифікація та систематизація документів за обов'язковими атрибутами.

**Апробація результатів дисертації.** Основні результати наукових досліджень неодноразово доповідалися на ряді міжнародних та всеукраїнських наукових конференцій, зокрема: the 4-th, 7-th, 8-th, 11-th International Conferences «Computer Science and Information Technologies» (Lviv, CSIT-2009, CSIT-2012, CSIT-2013, CSIT-2016); the XXI Polish-Ukrainian Conference “CAD in Machinery Design –

Implementation and Educational Problems” (Poland, Łańcut: CADMD-2013); Human. Computer. Communication: the 1st International Scientific Conference HCC-2013 (Lviv, 2013); V Міжнародна науково-практична конференція “Математика. Інформаційні технології. Освіта” (Луцьк-Світязь, 2016); II, III, IV, V Міжнародні наукові конференції “Інформація, комунікація, суспільство” (Львів-Славське, 2013, 2014, 2015, 2016); IV, V-th International Conferences of Young Scientists «Computer Science and Engineering» (Lviv, CSE-2010, CSE-2011). Про результати дисертаційних досліджень автор регулярно доповідав на наукових семінарах кафедри інформаційних систем та мереж Національного університету “Львівська політехніка” (2011 – 2016).

**Публікації.** За результатами дисертаційних досліджень опубліковано 25 наукових праць, серед яких 1 стаття у закордонному періодичному виданні, 9 статей у фахових наукових виданнях України, 15 публікацій у матеріалах наукових конференцій, серед яких 1 – у закордонному виданні, 1 – в наукометричній базі Scopus.

**Структура та обсяг роботи.** Дисертаційна робота складається зі вступу, чотирьох розділів, висновків та списку використаних джерел із 170 найменувань. Загальний обсяг дисертації 188 сторінок, з них 146 сторінок основного тексту, 3 додатки, 42 рисунки та 1 таблиця.

## **ОСНОВНИЙ ЗМІСТ РОБОТИ**

У вступі обґрунтовано актуальність теми, сформульовано мету та основні завдання досліджень, показано зв'язок із науковими програмами, планами, темами, розкрито наукову новизну. Розглянуто практичну цінність, реалізацію та впровадження результатів роботи. Наведено дані про особистий внесок здобувача, апробацію роботи та публікації.

У першому розділі подано результати аналізу тенденцій розвитку туристичної галузі в сучасних умовах, зокрема, особливості формування туристичної документації та використання відкритих веб-ресурсів. Проаналізовано особливості використання відкритих веб-ресурсів для формування туристичної документації, підходи до опису метаданих та структурування інформації в XML-технологіях, що дозволило побудувати інформаційну модель туристичної документації та інформаційного джерела з використанням індикативних ознак на основі лінгвістичних маркерів, а також обґрунтовано вибір методу оцінювання якості туристичної документації на основі стандарту ISO/IEC-25010, який є основою для розробки методів і процедур, що застосовуються під час формування туристичної документації.

**Туристична послуга (ТП)** – послуга суб'єкта туристичної діяльності щодо розміщення, харчування, транспортного, інформаційно-рекламного обслуговування, а також послуги закладів культури, спорту, побуту, розваг тощо, спрямована на задоволення потреб туристів

**Туристичний продукт** – попередньо розроблений комплекс туристичних послуг, який поєднує не менше ніж дві такі послуги, що реалізується або пропонується для реалізації за визначеною ціною

**Туристична документація (ТД)** – це технічна документація про туристичні послуги і/або продукти, яка необхідна для споживача туристичної послуги і/або продукту.

**Відкриті веб-ресурси (ВВР)** – ресурси WWW, які знаходяться у відкритому доступі для усіх користувачів Інтернет.

Аналіз особливостей ВВР показав, що для ефективного формування ТД потрібно стандартизувати опис метаданих і привести різні метадані до єдиного формату. Для цього використано стандарт Dublin Core Metadata Initiative (Дублінське ядро), що містить словник відповідних основних понять, а також встановлено, що підходи, які використовуються для структурування інформації на веб-ресурсах, зокрема, XML-засоби структурування інформаційного наповнення ТД, впливають на процедури формування ТД.

Для визначення якості туристичної документації розроблено критерії оцінювання показників якості згідно з стандартом ISO/IEC 25010, який полягає у застосуванні показників якості стандарту ISO/IEC-25010 до оцінювання якості складових ТД і мають відповідати таким вимогам: забезпечувати певну множину структурних елементів, які наповнюються інформацією про туристичну діяльність; бути зрозумілою для споживача; простою та зручною у використанні; не містити прогалин в наповненні структурних елементів; містити актуальну інформацію; задовільняти потреби споживача.

Аналіз метрик, що відповідають характеристикам/під-характеристиками якості ПЗ за стандартом ISO/IEC-25010, показує, що для оцінювання якості ТД придатними є такі характеристики/під-характеристики: функціональність (Functional suitability), придатність до використання (Operability), супроводжуваність (Maintainability).

У другому розділі наведено результати розробки інформаційної моделі туристичної документації, моделі інформаційних джерел формування туристичної документації, а також розроблено метод лінгвістичного співставлення атомарної ситуації з елементами туристичної документації.

Загальну структуру інформаційної моделі туристичної документації подано у вигляді ієрархії (дерева) чотирьох рівнів: 1-й рівень – туристична документація; 2-й рівень – туристичні послуги та відгуки споживачів; 3-й рівень – туристичні об'єкти і дії над ними, які стосуються ТП і відгуків; 4-й рівень – факти та події, які стосуються ТО і ДТО (рис. 1).

На основі загальної структури ТД (рис. .1) побудовано формальну модель ТД, яку подано у вигляді кортежа:

$$TD = \langle TS, TE \rangle, \quad (1)$$

де  $TD$  – туристична документація,  $TS$  – множина туристичних послуг,  $TE$  – множина відгуків споживачів туристичних послуг.

**Туристичний об'єкт (ТО)** – це сукупність фактів та подій, які цілісно відображають відомості про реальний об'єкт, над яким здійснюють певну туристичну діяльність і є цікавий для туристів.

**Дії над туристичним об'єктом (ДТО)** – це сукупність фактів та подій, які цілісно відображають відомості про реальну діяльність над відповідним реальним об'єктом. Структура ДТО аналогічна ТО.



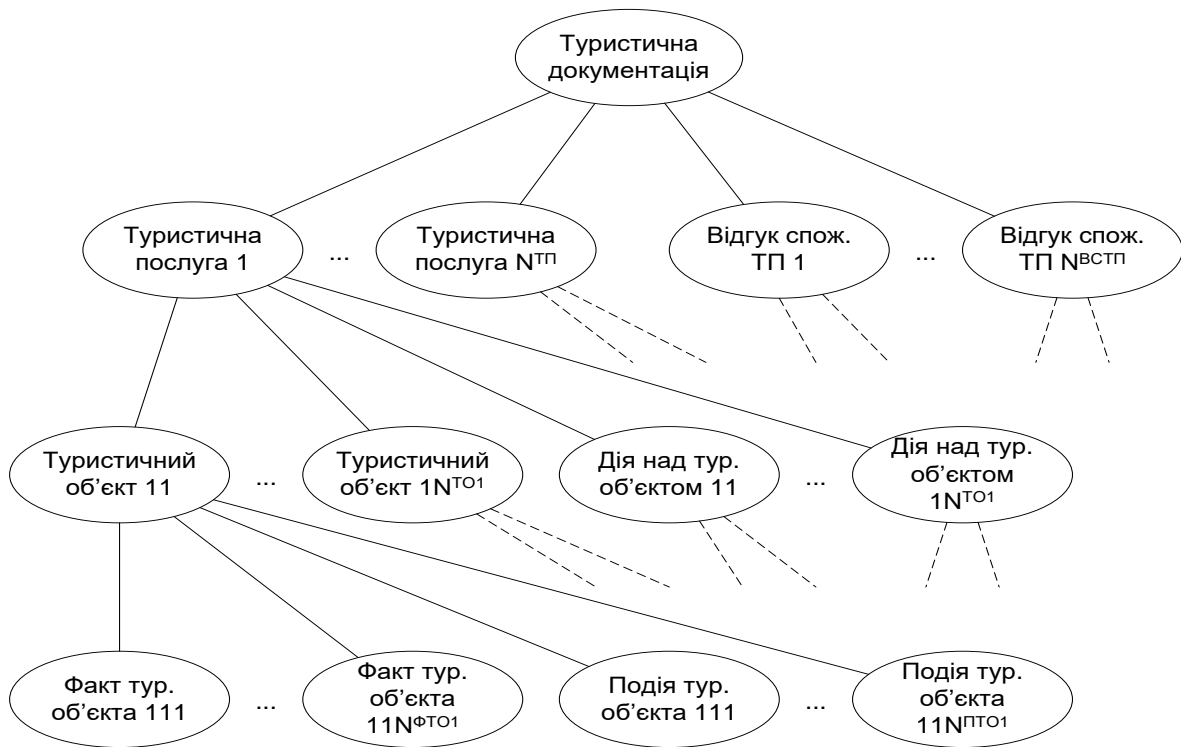


Рис.1. Загальна структура туристичної документації

Туристичну послугу подано у вигляді кортежа:

$$TS_i = \langle TSO_i, TSA_i \rangle, \quad (2)$$

де  $TSO_i$  – множина об'єктів ТП (тобто ТО),  $TSA_i$  – множина дій над об'єктами ТП (тобто ДТО).

**Відгук споживача** – це множина відгуків про ТО і про ДТО, що складатися з фактів і подій, які трапилися з цим споживачем і пов'язані з цим об'єктом чи дією.

Відгук споживачів ТП подамо у вигляді кортежа:

$$TE_i = \langle TEO_i, TEA_i \rangle,$$

де  $TEO_i$  – множина відгуків про ТО,  $TEA_i$  – множина відгуків про ДТО

Поняття відгуків споживачів ТП та їх складових формалізовано аналогічно до (1) та (2). Кожен ТО  $Object_i$  з множини усіх можливих об'єктів ТД – це кортеж такого вигляду:

$$Object_i = \langle ODefinition_i, OFact_i, OEvent_i \rangle,$$

де  $ODefinition_i$  – множина атрибутів ТО,  $OFact_i$  – множина фактів про ТО,  $OEvent_i$  – множина подій про ТО.

$$ODefinition_i = \langle MandatoryProperties(Object_i), Classifier(Object_i) \rangle,$$

де  $MandatoryProperties(...)$  – обов'язкові атрибути поняття як Ідентифікатор, Заголовок/Назва, Опис,  $Classifier(...)$  – множина термінів класифікаторів, до яких відноситься поняття.

**Факти та події** – це елементарні (атомарні) відомості про туристичні послуги і туристичні об'єкти, які відповідно не мають та мають прив'язку до часу. З фактів та подій складаються такі складові ТД як туристичні об'єкти, дії над туристичними

об'єктами, відгуки до туристичних об'єктів, відгуки до дій над туристичними об'єктами. З іншого боку кожні факти та події ТД становлять окремі документи, які фіксують відомості про ТО, ДТО, відгуки до ТО та ДТО на основі інформації, що міститься у відповідному ВВР.

На основі проведеного аналізу підходів до моделювання ВВР побудовано загальну модель інформаційного джерела ТД:

$$Source = Reduction(Site) = \langle MandatoryProperties(Site), Page, RPageRels \rangle,$$

де  $MandatoryProperties(Site)$  – обов'язкові атрибути сайту (адреса, назва, гасло тощо),  $Page = \{Page_i\}_{i=1}^{N^{(Page)}}$  – множина сторінок сайту, що є сукупністю певних атрибутів, інформаційних блоків, що визначаються множиною інформаційних дописів та співвідношень між цими блоками,  $N^{(Page)}$  – кількість сторінок сайту,  $RPageRels \subseteq PageRels$  – множина посилань між сторінками, яка забезпечує односторонню зв'язність від головної сторінки  $Page_1$ , і достатня для того, щоб цей граф не містив циклів.

Розроблено метод лінгвістичного співставлення *атомарної ситуації* (певної частини тексту допису, яка відображає відомості про окремі факти або події) з елементами ТД на основі лінгвістичних маркерів та модифікованих індикативних ознак.

*Маркер атомарної ситуації* – фраза, слово або частина слова, яка вказує (маркує) роль речення чи групи речень і використовується для визначення механізмів виявлення атомарної ситуації в дописах.

Введено такі типи маркерів:

1.  $ObjectMarker = \{ObjectMarker_i\}_{i=1}^{N^{(ObjectMarker)}}$  – множина маркерів ТО (такі фрази як 'Львів', 'Карпати', 'Ратуша' тощо),  $N^{(ObjectMarker)}$  – кількість маркерів ТО;
2.  $TermMarker = \{TermMarker_i\}_{i=1}^{N^{(TermMarker)}}$  – множина маркерів термінів класифікаторів ТД (такі фрази як 'харчування', 'проживання' тощо),  $N^{(TermMarker)}$  – кількість маркерів класифікаторів ТД;
3.  $ActionMarker = \{ActionMarker_i\}_{i=1}^{N^{(ActionMarker)}}$  – множина маркерів дій над ТО (такі фрази як 'відвідування', 'огляд', 'проходження' тощо),  $N^{(ActionMarker)}$  – кількість маркерів дій над ТО;
4.  $ExperienceMarker = \{ExperienceMarker_i\}_{i=1}^{N^{(ExperienceMarker)}}$  – множина маркерів відгуків споживачів (такі фрази як 'я був', 'бачив', 'проживав' тощо),  $N^{(ExperienceMarker)}$  – кількість маркерів відгуків споживачів;
5.  $TimeMarker = \{TimeMarker_i\}_{i=1}^{N^{(TimeMarker)}}$  – множина маркерів часових складових у дописах (час публікації, точні дати, такі фрази як 'вчора', 'місяць тому' тощо),  $N^{(ObjectMarker)}$  – кількість маркерів ТО;

*Індикативна ознака* – сукупність маркерів, яким притаманні певні спільні лінгвістичні особливості.

$$ObjectIndicator(Object_i) = \left\{ \left\{ Marker(Object_i)_j, \mu(Object_i)_j \right\} \right\}_{j=1}^{N(ObjectIndicator(Object_i))},$$

де  $\mu(Object_i)_j \in [0,1]$  – міра відповідності  $j$ -го маркера до  $i$ -го об'єкта (табл.1)

Якщо  $Object_p \in Object$  –  $p$ -й об'єкт ТД, то з урахуванням індикативної ознаки відповідність маркера ТО атомарної ситуації до об'єкта  $Object_p$  – це така величина:

$$ObjectAdequacy(Object_p, ObjectMarker(PostPart_{ijkl})) =$$

$$Rating(Page_i) * Trust(Author(Post_{ijk})) * \mu(Object_p, ObjectMarker(PostPart_{ijkl})),$$

де  $Rating(Page_i)$  – рейтинг  $i$ -ї веб-сторінки,  $Trust(Author(Post_{ijk}))$  – рівень довіри до автора допису. Для деякого  $q$  виконується:

$$\mu(Object_p, ObjectMarker(PostPart_{ijkl})) =$$

$$\begin{cases} \mu(Object_p)_q, & \text{якщо } ObjectMarker(PostPart_{ijkl}) = Marker(Object_p)_q; \\ 0, & \text{якщо } ObjectMarker(PostPart_{ijkl}) \neq Marker(Object_p)_q. \end{cases}$$

На відміну від маркерів ТО маркери відгуку споживача показують, чи можна вважати атомарну ситуацію дійсно думкою споживача про ТО.

Таблиця 1.

Приклад індикативної ознаки об'єкта туристичної документації 'Ратуша'

Маркер	Міра відповідності до об'єкта
'*Льв?в*Ратуш*'	1
'*Льв?в*веж*центр*'	0,7
'*Льв?в*міськ*веж*'	0,9
'*Льв?в*веж*міська*рад*'	0,95
'*культ*столиц*Україн*Ратуш*'	0,5
'*Ратуш*площ*Ринок*'	0,5
'*Ратуш*кам'ян*веж*'	0,4
'*Ратуш*кам'ян*веж* площ*Ринок*'	0,5
'*Льв?в*центр*'	0,3

Відповідність маркерів сукупності семантично-тотожних ситуацій визначається на основі сукупної міри відповідності атомарних ситуацій до елемента ТД:

$$\mu(Object_p, ObjectMarker(PostPart_{ijkl_1}), ObjectMarker(PostPart_{ijkl_2})) =$$

$$\mu(Object_p, ObjectMarker(PostPart_{ijkl_1})) +$$

$$(1 - \mu(Object_p, ObjectMarker(PostPart_{ijkl_1}))) \mu(Object_p, ObjectMarker(PostPart_{ijkl_2})).$$

Тоді відповідність маркерів ТО сукупності семантично тотожних атомарних ситуацій до об'єкта є наступна:

$$ObjectAdequacy \left( \begin{matrix} Object_p, ObjectMarker(PostPart_{ijkl_1}), \\ ObjectMarker(PostPart_{ijkl_2}) \end{matrix} \right) =$$

$$Rating(Page_i) * Trust(Author(Post_{ijk})) *$$

$$\mu(Object_p, ObjectMarker(PostPart_{ijkl_1}), ObjectMarker(PostPart_{ijkl_2})).$$

Аналогічно визначено інші відповідності маркерів сукупності семантично тотожних атомарних ситуацій.

У третьому розділі розроблено процедури комп'ютерно-лінгвістичного формування туристичної документації, які є необхідні для збору, опрацювання та збереження туристичної інформації, отриманої з ВВР.

Розроблено процедуру вибору ВВР для формування ТД, а саме визначено критерії відповідності ВВР до формування ТД (табл. 2), а також розроблено процедуру співставлення вмісту ВВР з елементами ТД на основі процедури розкладання дописів веб-сторінок на атомарні ситуації. Текст допису складається з атомарних фрагментів – синтаксичних структур, якими виступають або окремі речення, або марковані тегами фрагменти тексту.

Таблиця 2.

## Критерії відповідності ВВР до формування ТД

Авторитетність відкритих веб-ресурсів	Критерій авторитетності ВВР наступний: $Rating(Source) \geq \alpha^{(Rating(Source))},$ де $Rating(Source)$ – рейтинг ВВР, $\alpha^{(Rating(Source))} \in (0,1]$ – мінімально допустиме значення рейтингу ВВР.
Відповідність до тематики	Критерій відповідності до тематики наступний: $ExpertThemeEval(Source) \geq \alpha^{(ExpertThemeEval(Source))},$ де $\alpha^{(ExpertThemeEval(Source))} \in (0,1]$ – мінімально допустиме значення експертної оцінки відповідності ВВР до тематики ТД. Критерій відповідності ВВР за релевантністю до тематики наступний: $RelevanceThemeEval(Source) = 1 - \frac{N^{(RelevanceThemeEval(Source))} - 1}{N^{(RelevanceThemeEval)} - 1} \geq 0,$ де $N^{(RelevanceThemeEval(Source))}$ – позиція Веб-сторінки джерела в результаті виконання розширеного параметризованого пошукового запиту
Достовірність дописів	Критерій достовірності дописів ВВР наступний: $Trust(Author(Source)) \geq \alpha^{(Trust(Author(Source)))},$ де $\alpha^{(Trust(Author(Source)))} \in (0,1]$ – мінімально допустиме значення сумарного рівня довіри до авторів ВВР
Актуальність дописів	Актуальність дописів визначено часом їхньої публікації допису: $Actuality(Post_{ijk}) = \begin{cases} \frac{(T_{now} - T_0) - (T_{now} - PostTime(Post_{ijk}))}{T_{now} - T_0}, & PostTime(Post_{ijk}) \geq T_0, \\ 0, & PostTime(Post_{ijk}) < T_0 \end{cases}$ де $PostTime(Post_{ijk})$ – час публікації, $T_0$ – момент часу початку відліку актуальності, $T_{now}$ – поточний момент часу.

Алгоритми розкладання дописів на атомарні ситуації здійснено за двома методами: методом декомпозиції (від тексту допису відкидаються синтаксичні структури доти, доки результуючий фрагмент не буде стосуватися одного маркера ТО); методом синтезу (синтаксична структура допису, яка стосується одного

маркера ТО, доповнюється наступними синтаксичними структурами доти, поки результуючий фрагмент буде стосуватися одного ТО).

Для виявлення недоліків, які можуть виникнути в результаті формування туристичної документації розроблено процедуру виявлення прогалін. *Прогалина* в ТД – це певна частина ТД, яка не наповнена інформацією про факти чи події і є двох видів: структурні прогаліни в ТД та прогаліни у вмісті ТД.

*Структурна прогалина* – множина однотипних елементів ТД містить менше елементів, ніж відповідна множина цих елементів у вимогах до ТД:

$$Element^{(Gap)} = Element^{(Requirement)} \setminus Element^{(TD)} \neq \emptyset,$$

де  $Element^{(TD)}$  – множина елементів в ТД,  $Element^{(Requirement)}$  – множина елементів у вимогах до ТД

*Прогалина вмісту* визначається тоді, коли для структурного елемента ТД не усі необхідні факти та події з ним пов'язані:

$$ElementDefiniton^{(Gap)}(Term_{ij}) = \\ ElementDefiniton^{(Expected)}(Term_{ij}) \setminus ElementDefiniton^{(TD)}(Term_{ij}) \neq \emptyset,$$

де  $Term_{ij}$  – термін класифікатора,  $ElementDefiniton^{(TD)}(Term_{ij})$  – множина усіх фактів визначення ТО і ДТО, пов'язаних з цим терміном в ТД,  $ElementDefiniton^{(Expected)}(Term_{ij})$  – множина усіх фактів визначення ТО і ДТО.

Для усунення некоректної інформації у ТД розроблено процедуру виявлення надлишкової інформації в ТД та процедуру усунення неактуальної інформації в ТД. Процедура виявлення надлишкової інформації полягає у виявленні двох подібних елементів ТД (фактів чи подій):

$$Similarity(ElementFact_1, ElementFact_2) = \\ Similarity(Description(ElementFact_1), Description(ElementFact_2)),$$

де  $Description(ElementFact)$  – один з обов'язкових атрибутів факту, який відображає його текстовий опис,  $Description(ElementEvent)$  – один з обов'язкових атрибутів події, який відображає її текстовий опис.

При виявленні двох подібних елементів (фактів чи подій) необхідно здійснити процедуру усунення надлишковості шляхом об'єднання елементів.

Процедура усунення неактуальної інформації базується на перевірці фактів чи подій на предмет входження їх у термін актуальності за таким критерієм:

$$T_{NOW} - Time(ElementEvent)_i \leq T^{(Actuality)},$$

де  $T_{NOW}$  – поточний момент часу,  $T^{(Actuality)}$  – термін (часовий відрізок) актуальності.

У **четвертому розділі** розроблено показники якості туристичної документації, розроблено систему комп'ютерно-лінгвістичного формування туристичної документації, зокрема, розроблено архітектуру системи автоматизованого формування туристичної документації, а також наведено результати практичного впровадження запропонованих методів та засобів.

Розроблено метод оцінювання якості туристичної документації, за такими показниками відповідно до стандарту ISO/IEC 25010 (табл. 3): функціональні (відповідність / appropriateness), показники якості придатності до використання (відповідність розпізнавання / appropriateness recognisability, рівень допомоги / helpfulness, привабливість / attractiveness), показники якості супроводжуваності туристичної документації (модульність / modularity, стійкість до змін / modification stability).

Таблиця 3.

## Показники якості туристичної документації

№ п/п	Показник якості	Формула
1.	Відповідність	$Quality^{(FS)} = 1 - A^{(FS)} / B^{(FS)}$ , де $A^{(FS)}$ – множина структурних елементів ТД з структурними прогалинами, $B^{(FS)}$ – кількість елементів множини одного з структурних елементів ТД за вимогами;
2.	Відповідність розпізнавання	$Quality^{(O.AR)} = A^{(O.AR)} / B^{(O.AR)}$ , де $A^{(O.AR)}$ – кількість структурних елементів ТД, що містять прогалини вмісту, $B^{(O.AR)}$ – загальна кількість наявних структурних елементів ТД;
3.	Рівень допомоги	$Quality^{(O.H)} = A^{(O.H)} / B^{(O.H)}$ , де $A^{(O.H)}$ – кількість ТО або ДТО, до яких є відгуки споживачів, і які не мають прогалин вмісту, $B^{(O.H)}$ – кількість елементів ТО або ДТО, до яких є відгуки споживачів;
4.	Привабливість	$Quality^{(O.A)} = A^{(O.A)} / B^{(O.A)}$ , де $A^{(O.A)}$ – кількість елементів ТО або ДТО, до яких є відгуки споживачів, $B^{(O.A)}$ – кількість елементів множини наявних ТО або ДТО в ТД;
5.	Модульність	$Quality^{(M.M)} = A^{(M.M)} / B^{(M.M)}$ , де $A^{(M.M)}$ – кількість ТП або відгуків споживачів, які містять від $N_{\min}^{(Element)}$ до $2N_{\min}^{(Element)}$ ТО і ДТО, $B^{(O.A)}$ – загальна кількість ТП або відгуків споживачів;
6.	Стійкість до змін	$Quality^{(M.MS)} = \frac{N_2^{(M.M)} / T_2^{(M.M)}}{N_1^{(M.M)} / T_1^{(M.M)}}$ , де $N_1^{(M.M)}$ – кількість усіх невідповідностей, які зустрічаються в попередніх показниках якості до зміни ТД, $N_2^{(M.M)}$ – кількість усіх невідповідностей, які зустрічаються в попередніх показниках якості після зміни ТД, $T_1^{(M.M)}$ – інтервал часу, протягом якого виявлялися невідповідності до зміни ТД, $T_2^{(M.M)}$ – інтервал часу, протягом якого виявлялися невідповідності після зміни ТД.

На основі розроблених процедур формування ТД розроблено архітектуру комплексної інформаційної системи (ІС) авоматизованого формування туристичної документації, яка забезпечує ефективне керування інформацією для потреб туристичної установи та для оперативного задоволення запитів користувачів, а також забезпечує формування якісної ТД на основі ВВР відповідно до розроблених показників якості і функціонує на трьох архітектурних рівнях (рис. 2): рівень

локальної ІС; рівень агентів формування ТД; рівень зовнішніх інформаційних сервісів та ресурсів.

Комп'ютерно-лінгвістичну обробку застосовано на другому рівні за допомогою агента вибору відповідних джерел інформації, агента сканування веб-сторінок і розкладання на атомарні ситуації та агента співставлення атомарних ситуацій з елементами ТД.

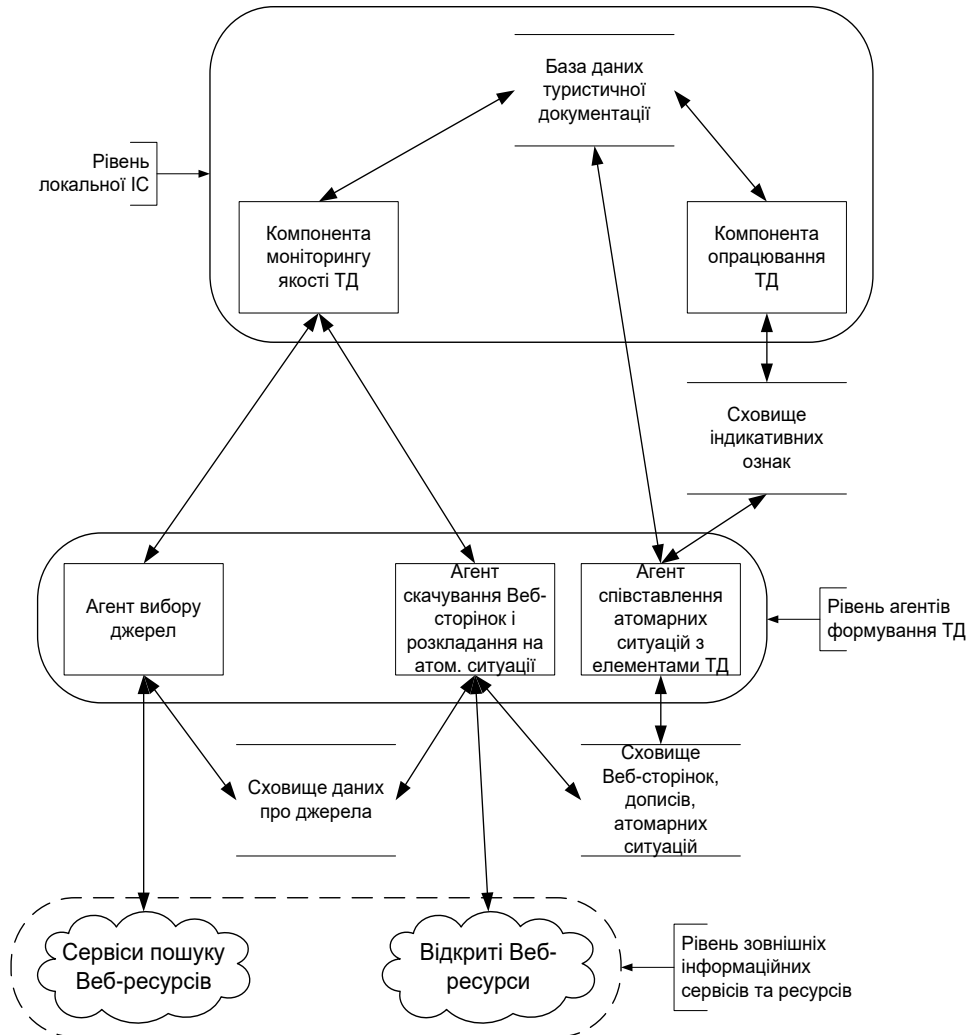


Рис. 2. Архітектура комплексної інформаційної системи автоматизованого формування туристичної документації

На рис. 3 – рис. 5 зображена динаміка стабілізації рівня ТД в агенствах «Тур де Львів» та «Королівські подорожі» після формування і підтримки ТД в актуальному стані. Встановлено, що прийнятний рівень залежить від показника якості туристичної документації або групи показників.

Аналіз впровадження результатів досліджень на практиці підтвердив ефективність розроблених методів і засобів формування туристичної документації за допомогою комп'ютерно-лінгвістичного аналізу інформаційного наповнення відкритих веб-ресурсів, а також підтвердив істотний ріст (в 1,5-3 рази) кількості структурних елементів з можливістю їхньої належної підтримки з показниками якості інформації про них від 0,6 до 1.

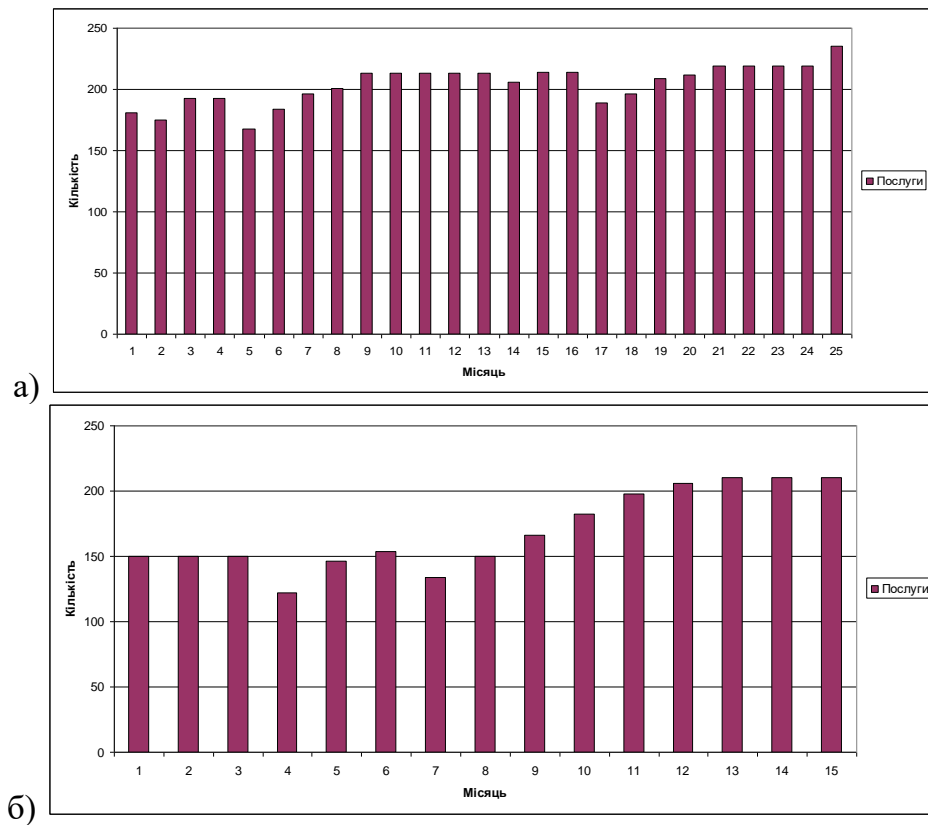


Рис. 3. Динаміка зміни кількості послуг у туристичній документації агенства «Королівські подорожі» (а) та «Тур де Львів» (б)

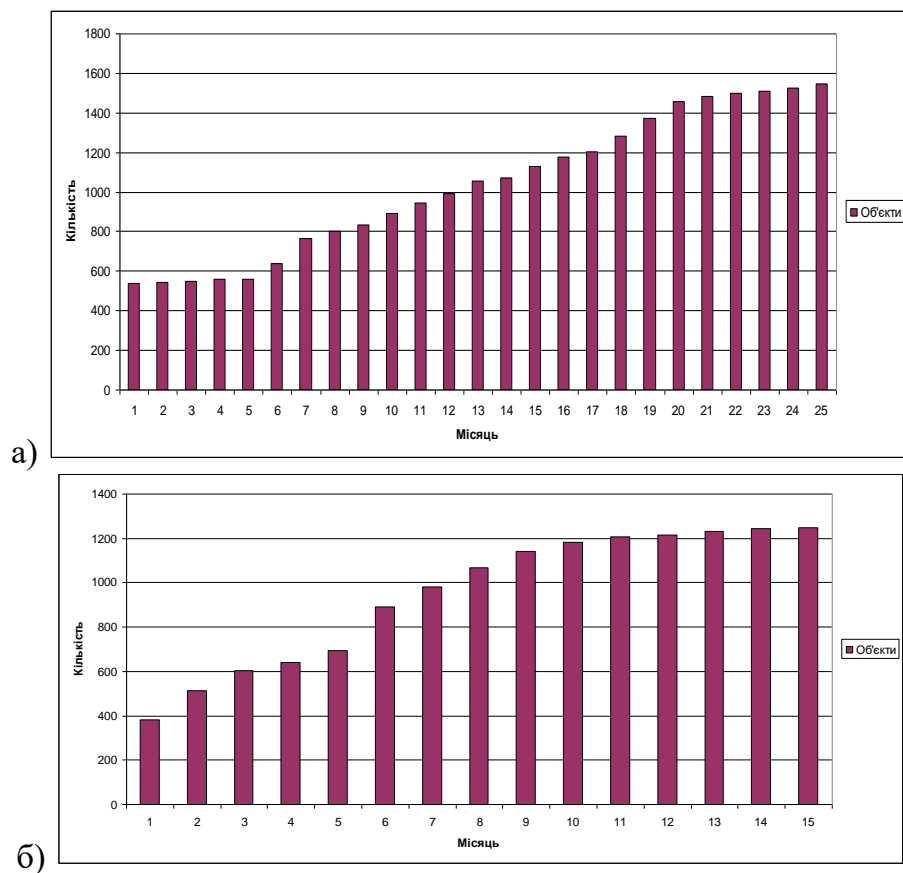


Рис. 4. Динаміка зміни кількості об'єктів у туристичній документації агенства «Королівські подорожі» (а) та «Тур де Львів» (б)



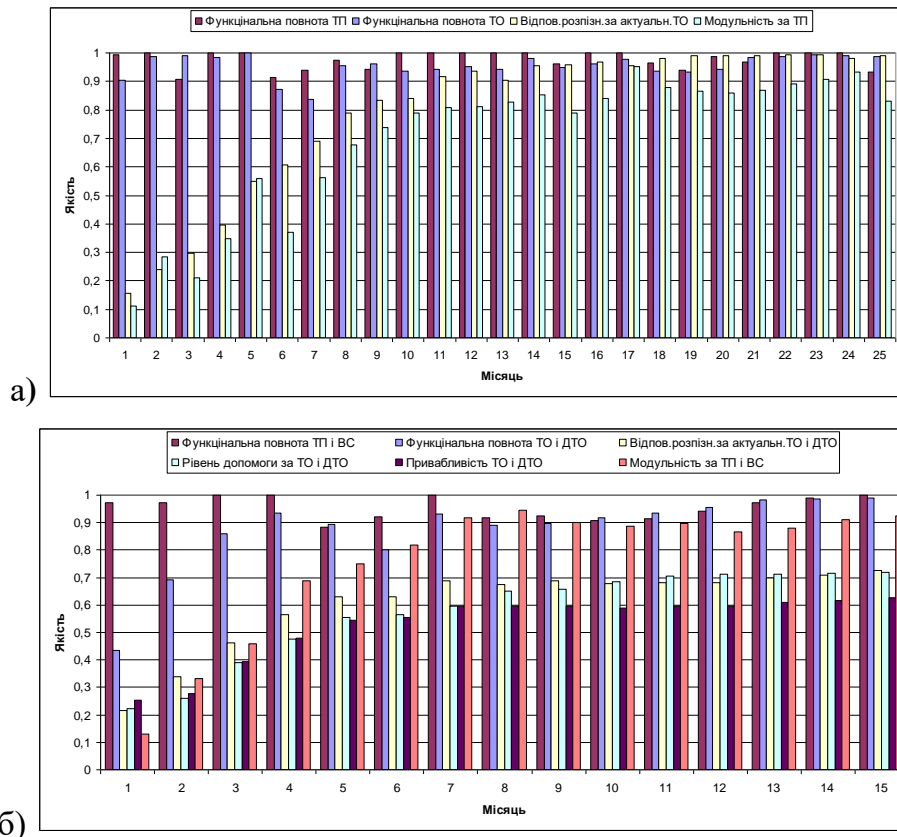


Рис. 5 Динаміка зміни показників якості туристичної документації агенства «Королівські подорожі» (а) та «Тур де Львів» (б)

## ОСНОВНІ РЕЗУЛЬТАТИ ТА ВИСНОВКИ

У дисертаційній роботі розв'язано важливе наукове завдання розроблення методів і засобів формування туристичної документації шляхом комп'ютерно-лінгвістичного аналізу інформаційного наповнення відкритих веб-ресурсів.

Основними науковими та практичними результатами є:

1. Визначено основні підходи до формування туристичної документації на основі відкритих веб-ресурсів шляхом аналізу методів структурування та інтеграції інформаційного наповнення веб-ресурсів, що стало основою для моделей туристичної документації та інформаційних джерел.

2. Уперше розроблено формальну модель туристичної документації на основі множин туристичних послуг, відгуків споживачів туристичних послуг та класифікаторів з урахуванням відповідних індикативних ознак їх елементів, що стало основою для розробки процедур комп'ютерно-лінгвістичного формування туристичної документації і показників її якості.

3. Удосконалено формальну модель інформаційного джерела для формування туристичної документації шляхом виокремлення блоків і атомарних ситуації туристичної інформації на веб-сторінках, що стало основою для побудови процедури вибору відкритих веб-ресурсів, процедур лінгвістичного співставлення вмісту відкритих веб-ресурсів з елементами туристичної документації.

4. Набув подальшого розвитку метод лінгвістичного виявлення інформації у веб-ресурсах шляхом співставлення атомарних ситуацій з елементами туристичної

документації на основі модифікованих індикативних ознак, що дозволило побудувати відповідні процедури лінгвістичного співставлення вмісту відкритих веб-ресурсів з елементами туристичної документації.

5. Розроблено процедури комп'ютеро-лінгвістичного формування туристичної документації на основі співставлення вмісту відкритих веб-ресурсів з елементами за допомогою індикативних ознак, а також процедури виявлення прогалин та некоректної інформації у туристичній документації на основі аналізу її структури та вмісту, які дозволяють покращити показники якості туристичної документації і є необхідними для побудови системи автоматизованого формування туристичної документації.

6. Уперше розроблено метод оцінювання якості туристичної документації на основі функціональних показників якості, показників якості придатності до використання та показників якості супроводжуваності стандарту ISO/IEC-25010, що забезпечило практичне оцінювання результатів виконання процедур формування та перевірки туристичної документації.

7. Розроблено архітектуру та функціональну структуру системи комп'ютерно-лінгвістичного формування туристичної документації на основі процедур формування та показників якості туристичної документації, що на практиці дозволило істотно збільшити (в 1,5-3 рази) кількість структурних елементів з можливістю їхньої належної підтримки з показниками якості інформації про них від 0,6 до 1.

## **СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ**

1. Жежнич П. Особливості формування електронної документації в сфері туризму / П. Жежнич, О. Сопрунюк // Інформаційні системи та мережі : Вісник Національного університету "Львівська політехніка". – Львів, 2013. – № 770. – С. 101-108.
2. Markiv O. O. Peculiarities of the dublin core use in the formation process of tourism documentation / Markiv O. O. // Комп'ютерні науки та інформаційні технології : Вісник Національного університету "Львівська політехніка". – Львів : Видавництво Львівської політехніки, 2016. – № 843. – С. 104-107.
3. Tkachenko S. Efficiency enhancement of optimal reduction method by strengthening parallelism of structural models formation / S. Tkachenko, O. Soprnyuk, V. Tkachenko, I. Solomko // Machine Dynamics Research. – Poland, 2013. – Vol. 3, No. 2. – P. 85-90.
4. Сопрунюк О.О. Метод оцінювання якості туристичної документації / Сопрунюк О.О. // Інформаційні системи та мережі : Вісник Національного університету "Львівська політехніка". – Львів, 2014. – № 805. – С. 333-343.
5. Жежнич П. Консолідація відкритих інформаційних ресурсів в туристичній сфері / П. Жежнич, О. Сопрунюк // Комп'ютерні науки та інформаційні технології : Вісник Національного університету "Львівська політехніка". – Львів, 2013. – № 771. – С. 3-11.
6. Жежнич П. Основні підходи до моніторингу та аналізу якості туристичної документації / П. Жежнич, О. Сопрунюк // Інформаційні системи та мережі :

- Вісник Національного університету “Львівська політехніка”. – Львів, 2012. – № 743. – С. 190-199.
7. Пелецишин А. Роль прогалін у документації в процесах виготовлення та експлуатації продукції / А. Пелецишин, О. Сопрунюк, О. Тимовчак-Максимець // Комп’ютерні науки та інформаційні технології: Вісник Національного університету “Львівська політехніка”. – Львів, 2011. – № 719. – С. 17-22.
  8. Tkachenko S. Models of optimal reduction in decomposition problems solution and structures modeling / S. Tkachenko, O. Soprunyuk // Комп’ютерні системи проектування. Теорія і практика: Вісник Національного університету “Львівська політехніка”. – Львів, 2013. – № 777. – Р. 12-15.
  9. Жежнич П. Особливості формування інформаційного забезпечення туристичної діяльності / П. Жежнич, О. Сопрунюк // Інформаційні системи та мережі: Вісник Національного університету “Львівська політехніка”. – Львів, 2014. – № 783. – С. 336-343.
  10. Жежнич П. Аналіз сучасних засобів автоматизації роботи з електронною документацією / П. Жежнич, О. Сопрунюк, О. Марчик // Комп’ютерні науки та інформаційні технології: Вісник Національного університету “Львівська політехніка”. – Львів, 2012. – № 732. – С. 395-398.
  11. Zhezhnych P. Analysis of the tourism documentation quality improvement / P. Zhezhnych, O. Soprunyuk // Materials of the 7th International Scientific and Technical Conference “Computer Science and Information Technologies (CSIT’2012)”. – Lviv : Publishing house “Vezha and Co”, 2012. – Р. 34-36.
  12. Markiv O. Graphical presentation of tourism documentation structure / Markiv O. // Міжнародна наукова конференція «Інформація. Комунікація. Суспільство» (ICS’2016). – Славське-Львів, 2016. – Р. 216-217.
  13. Марків О. Формування туристичної документації на основі виявлення та заповнення прогалін / Марків О. // Інформація, комунікація, суспільство (ІКС-2015): матеріали I Міжнародної наукової конференції. – Славське-Львів: Видавництво НУ «Львівська політехніка», 2015. – С. 230-231.
  14. Tkachenko S. Fast formation of binary trees as parallel (optimal) reduction process models / S. Tkachenko, O. Soprunyuk, V. Tkachenko, N. Yatsyk // CAD in Machinery Design – Implementation and Educational Problems: materials of the XXI Polish-Ukrainian Conference – CADMD 2013. – Poland, Łańcut: Warsaw University of Technology, 2013. – Р. 49.
  15. Zhezhnych P. Peculiarities of the open web-resources use in the formation process of tourism documentation / P. Zhezhnych, O. Soprunyuk // VI Міжнародна конференція молодих вчених “Computer Sciences and Information Technologies” (“Комп’ютерні науки та інженерія”) (CSE’2013). – Львів, 2013. – С. 218-219.
  16. Zhezhnych P. Importance of the tourism information quality monitoring / P. Zhezhnych, O. Soprunyuk // Human. Computer. Communication: materials of the 1st International Scientific Conference (HCC-2013). – Lviv: Lviv Polytechnic National University publisher, 2013. – Р. 39.
  17. Zhezhnych P. Formation of the consolidated data resource of tourism information / P. Zhezhnych, O. Soprunyuk // VIII Міжнародна науково-технічна конференція

- “Computer Sciences and Information Technologies” (“Комп’ютерні науки та інформаційні технології”) (CSIT’2013). – 2013. – С. 194-195.
18. Zhezhnych P. Justification of use expedience of tourism data consolidation / P. Zhezhnych, O. Soprunyuk // Міжнародна наукова конференція «Інформація. Комунікація. Суспільство» (ICS’2014). – Славське-Львів, 2014. – С. 174-175.
  19. Zhezhnych P. Formation peculiarities of tourism documentation / P. Zhezhnych, O. Soprunyuk // Materials of the Second International Academic Conference “Information. Communication. Society” (IKS-2013). – Lviv : Lviv Polytechnic National University publisher, 2013. – P. 194-196.
  20. Melnykova N. Semantic approach to personalization of medical data / Melnykova N., Markiv O. // Матеріали XI Міжнародної науково-технічної конференції “Комп’ютерні науки та інформаційні технології” (CSIT’2016). – Львів : Видавництво Львівської політехніки, 2016. – С. 59-62.
  21. Andreychuk N. I. The contrastive analysis of the notion of Homo Socialis in English and Ukrainian / N. I. Andreychuk, O. Soprunyuk // Матеріали IV Міжнародної науково-технічної конференції “Computer Sciences and Information Technologies” (“Комп’ютерні науки та інформаційні технології”) (CSIT’09). – Львів : Видавництво ПП “Вежа і Ко”, 2009. – P. 465-468.
  22. Бойко Н. І. Приклад використання систем масового обслуговування в організації роботи web-систем / Н. І. Бойко, О. О. Марків // Матеріали V Міжнародної науково-практичної конференції “Математика. Інформаційні технології. Освіта”. – Луцьк, 2016. – С. 12-14.
  23. Кісь Я. П. Інтернет-меми як засіб розповсюдження інформації / Я. П. Кісь, О. О. Сопрунюк // Матеріали IV Міжнародної конференції молодих вчених “Комп’ютерні науки та інженерія (CSE-2010)”. – Львів, 2010. – С. 396-397.
  24. Жежнич П. Основні підходи до організації електронного організаційно-розпорядчого документообігу / П. Жежнич, О. Марчик, О. Сопрунюк // Матеріали I Міжнародної наукової конференції “Інформація, комунікація, суспільство (ІКС-2012)”. – Славське-Львів : Видавництво НУ «Львівська політехніка», 2012. – С. 34-36.
  25. Soprunyuk O. Lexical representation of human statuses in Tudor times: Tudor court titles / O. Soprunyuk // Матеріали IV Міжнародної конференції молодих вчених “Комп’ютерні науки та інженерія” (CSE-2011). – Львів : Видавництво НУ «Львівська політехніка», 2011. – С. 110-111.

## АНОТАЦІЇ

**Марків О. О. Лінгвістичне забезпечення формування туристичної документації на основі відкритих веб-ресурсів.** – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 10.02.21 – «Структурна, прикладна і математична лінгвістика». – Національний університет «Львівська політехніка» МОН України, Львів, 2017.

У дисертації вирішено науково-прикладне завдання розроблення методів і засобів комп’ютерно-лінгвістичного аналізу інформаційного наповнення відкритих веб-ресурсів для формування туристичної документації. Для розв’язання цього

завдання розроблено формальну модель туристичної документації та інформаційного джерела для формування туристичної документації.

Розроблено метод лінгвістичного співставлення атомарної ситуації з елементами туристичної документації на основі модифікованих індикативних ознак, що дало можливість побудувати відповідні процедури лінгвістичного співставлення вмісту відкритих веб-ресурсів з елементами туристичної документації, за допомогою лінгвістичних маркерів та індикативних ознак, а також процедури виявлення прогалин у туристичній документації, які дають можливість покращити показники якості туристичної документації.

Розроблено метод оцінювання якості туристичної документації на основі показників якості за стандартом ISO/IEC-25010. Також розроблено архітектуру інформаційної системи туристичної документації на основі процедур формування туристичної документації та показників якості туристичної документації.

Ключові слова: туристична документація, відкритий веб-ресурс, лінгвістичний маркер, індикативна ознака, якість туристичної документації.

**Маркив О. А. Лингвистическое обеспечение формирования туристической документации на основе открытых веб-ресурсов.** – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 10.02.21 – «Структурная, прикладная и математическая лингвистика». – Национальный университет «Львівська політехніка» Министерства образования и науки Украины, Львов, 2017.

В диссертации решено научно-прикладное задание разработки методов и средств компьютерно-лингвистического анализа информационного наполнения открытых веб-ресурсов для формирования туристической документации. Для решения этого задания разработано формальную модель туристической документации и информационного источника для формирования туристической документации.

Разработан метод лингвистического сопоставления атомарной ситуации с элементами туристической документации на основе модифицированных индикативных признаков, что позволило построить соответствующие процедуры лингвистического сопоставления содержимого открытых веб-ресурсов с элементами туристической документации, с помощью лингвистических маркеров и индикативных признаков, а также процедуры выявления пробелов в туристической документации, которые позволяют улучшить показатели качества туристической документации.

Разработан метод оценивания качества туристической документации на основе показателей качества согласно стандарта ISO/IEC-25010. Также разработана архитектура информационной системы туристической документации на основе процедур формирования туристической документации и показателей качества туристической документации.

Ключевые слова: туристическая документация, открытый веб-ресурс, лингвистический маркер, индикативный признак, качество туристической документации.

**Markiv O. O. Linguistic support of tourism documentation formation based on open web-sources.** – On the manuscript.

Thesis for a candidate's degree in technical sciences in speciality 10.02.21 – «Structural, applied and mathematical linguistics». – Lviv Polytechnic National University, Ministry of Education and Science of Ukraine, Lviv, 2017.

In the thesis scientific and practical task of methods and means development of computer-linguistic analysis of the open web-sources content for the tourism documentation formation process is solved. In order to solve this problem formal model of tourism documentation based on tourism services and tourism consumer experience sets and formal model of information web-source for the formation of tourism documentation based on web-pages information blocks and atomic situations are developed.

As the result the procedures of computer-linguistic tourism documentation formation based on linguistic comparison of open web-sources content with the elements of tourism documentation based on modified indicative characteristics and linguistic markers are created. Also procedures of identifying gaps and incorrect information in tourism documentation based on its structure and content analysis are developed. This allow to improve tourism documentation quality and is important for automatic system creation of tourism documentation creation.

Method of tourism documentation quality evaluation based on quality indices according to ISO/IEC-25010 (appropriateness, appropriateness recognisability, helpfulness, attractiveness, modularity, modification stability) is developed. Defined tourism documentation indices and procedures of tourism documentation creation allow to develop the architecture of the tourism documentation information system on three levels: local information system, agents of tourism documentation formation process and external information services and sources. It allows effectively to process tourism information and create quality tourism documentation based on open web-sources.

Analysis of research results has justified effectiveness of the developed methods and means of tourism documentation formation based on computer-linguistic analysis of open web-sources information content. It also presents the increase of tourism documentation structural elements quantity with theirs appropriate maintenance and quality indices from 0,6 to 1.

Key words: tourism documentation, open web-source, linguistic marker, indicative characteristics, tourism documentation quality.

Підписано до друку 01.03.2017 р.  
Формат 60×90 1/16. Папір офсетний.  
Друк на різнографі. Умовн. друк. арк. 1,5. Обл.-видав. арк. 0,89.  
Тираж 100 прим. Зам. 170225.

Поліграфічний центр  
Видавництва Національного університету "Львівська політехніка"  
вул. Ф.Колесси, 4, 79013, Львів  
*Ресстраційне свідоцтво серії ДК № 4459 від 27.12.2012 р.*