

УДК 004.8:004.912:001.811

Г.А. ДОБРОВОЛЬСЬКИЙ, Н.Г. КЕБЕРЛЕ, П.П. ПРОХОРЕНКО  
Запорізький національний університет**ЗАСТОСУВАННЯ МЕТОДІВ ПОБУДОВИ ТА АНАЛІЗУ МЕРЕЖІ  
ЦИТУВАННЯ ДЛЯ ПІДГОТОВКИ БІБЛІОГРАФІЧНОГО ПОКАЖЧИКА З  
МЕТОДІВ ПЕРСОНАЛІЗАЦІЇ СИСТЕМ КОМП'ЮТЕРНОГО НАВЧАННЯ  
ВИМОВИ**

*Розглядається застосування методу та інформаційної технології побудови і аналізу мережі цитування наукових публікацій з метою виявлення найбільш значущих статей в спеціальній області - методи персоналізації систем комп'ютерного навчання вимови. Технологія комбінує контрольований метод снігової кулі і аналіз головних шляхів. Розмір снігової кулі регулюється за допомогою ймовірнісної тематичної моделі і симетричної дивергенції Кульбака-Лейблера для відбору схожих статей. Головні шляхи в мережі цитування відображають розвиток знань у дослідженій області.*

*Ключові слова: аналіз текстів, короткі тексти, тематичне моделювання, метод головних компонент, розріджена симетрична невід'ємна матрична факторизація, мережа цитування, аналіз головних шляхів, метод снігової кулі, порушення вимови, адаптація до особливостей вимови, оцінка вимови.*

Г.А. ДОБРОВОЛЬСЬКИЙ, Н.Г. КЕБЕРЛЕ, П.П. ПРОХОРЕНКО  
Запорізький національний університет**ПРИМЕНЕНИЕ МЕТОДОВ ПОСТРОЕНИЯ И АНАЛИЗА СЕТИ ЦИТИРОВАНИЯ  
ДЛЯ ПОДГОТОВКИ БИБЛИОГРАФИЧЕСКОГО УКАЗАТЕЛЯ ПО МЕТОДАМ  
ПЕРСОНАЛИЗАЦИИ СИСТЕМ КОМПЬЮТЕРНОГО ОБУЧЕНИЯ ПРОИЗНОШЕНИЮ**

*Рассматривается применение метода и информационной технологии построения и анализа сети цитирования научных публикаций с целью обнаружения наиболее значимых статей в специальной области - методы персонализации систем компьютерного обучения произношению. Технология комбинирует контролируемый метод снежного шара и анализ главных путей. Размер снежного шара регулируется с помощью вероятностной тематической модели и симметричной дивергенции Кульбака-Лейблера для отбора похожих статей. Главные пути в сети цитирования отражают развитие знаний в исследованной области.*

*Ключевые слова: анализ текстов, короткие тексты, тематическое моделирование, метод главных компонент, разреженная симметричная неотрицательная матричная факторизация, сеть цитирования, анализ главных путей, метод снежного шара, нарушения речи, адаптация к особенностям произношения, оценка произношения.*

Н. DOBROVOLSKYI, N. KEBERLE, P. PROKHORENKO  
Zaporizhzhya National University**CITATION NETWORK CONSTRUCTION AND ANALYSIS METHOD APPLICATION TO  
BIBLIOGRAPHIC INDEX PREPARATION FOR PERSONALIZATION METHODS IN COMPUTER-  
AIDED PRONUNCIATION TRAINING**

*The paper presents an application of the method and correspondent information technology for construction and analysis of a citation network of scientific publications aimed at detecting the seminal papers in the particular domain of research - personalization methods in computer-aided pronunciation training.*

*The method relies on an initial set of seed papers arbitrarily selected by the domain(s) experts, adopts and combines probabilistic topic model, greedy restricted snowball sampling, and path analysis of the collected citation network. The topic model is built on the base of word-word co-occurrence probability with combination of sparse symmetric nonnegative matrix factorization and principal component approximation. Greedy restricted snowball sampling size is regulated with probabilistic topic model and symmetrized Kullback-Leibler divergence as measure of publication similarity. Path analysis allows further selection of the most important paths in the citation network and, hence, evaluation of the scientific value of each publication within the domain of interest. The method is robust with respect to the variations of initial set of seed papers, it preserves the scale-free property of a citation network, the restricted snowball sampling provides saturation of the publications set and reliable semantic distance between publications.*

*The research domain selected to apply the method lies at intersection of two scientific domains - speech disorders and computer-assisted pronunciation training, and as such it does not have a single seminal paper or*

*review to read. The constructed citation network allows a scientist willing to get acquainted with the domain to have a reading plan consisting of the most relevant, important and recognized publications.*

*The method applied can be used when a research domain is at intersection of several scientific domains, when a scientist joins a research field and does not have an experience in it, when a research domain is actively evolving and has been experiencing explosive growth in the number of publications.*

*Keywords: text mining, short text document, topic modelling, principal component analysis, sparse symmetric nonnegative matrix factorization, citation network, main path analysis, snowball sampling, speech disorder, speaker adaptation, pronunciation assessment.*

### **Постановка проблеми**

Автоматична оцінка якості вимови є важливою частиною комп'ютерних систем навчання вимові. Але у більшості випадків такі системи налаштовані на сприйняття мови типового учня, і чим більше відрізняється голос від ідеального, тим більшою є систематична помилка в оцінках. На практиці такі системи є занадто чутливими до індивідуальних особливостей голосу, що особливо помітно у випадках, якщо людині притаманні картавість, шепелявість, нерозбірлива вимова [1]. Серйозні захворювання, такі як дизартрія, зазвичай поєднують декілька особливостей вимови. Один із шляхів покращення точності оцінки є врахування персональних особливостей вимови автоматичними системами оцінки вимови. Дослідження [2,3] показало, що кожному окремому учню притаманно значно менше помилок вимови, ніж якщо брати усі можливі помилки вимови усіх учнів. Тому підвищення уваги автоматичної системи до персонального набору помилок набагато підвищує точність оцінки вимови. Дослідження такого типу відноситься до перетину кількох дисциплін: комп'ютеризовані системи оцінки вимови (САРТ), логопедія, обробка аудіо сигналів. Очевидно, не існує бібліографічних показників, що включали б всі значущі дослідження з усіх названих дисциплін, тому створення такого показника є актуальною задачею.

Якість аналізу існуючих публікацій у заданій області досліджень - проблема добре відома кожному науковцю. У процесі підготовки аналізу потрібно знайти публікації, які містять всі важливі наукові результати, виділити серед них головні ідеї та прослідкувати їх еволюцію. Але якісний підбір літератури стикається із перепонами: відсутність досвіду у заданій області досліджень, міждисциплінарні дослідження, велика кількість публікацій, якщо область знань активно розвивається.

Існують різноманітні способи автоматичного аналізу існуючих колекцій наукових публікацій: вивчення співавторства [4, 5], дослідження ключових слів та тем [6], пошук та аналіз термінології [7], обчислення різних статистичних показників [8], аналіз цитування [9, 8, 10]. Згадані методи успішно застосовуються до існуючих, ретельно зібраних та очищених даних, але збір нових репрезентативних колекцій на задану тему залишається складним та актуальним завданням із кількох причин. По-перше, існуючі колекції наукових статей не перекривають всі відомі напрямки досліджень, по-друге, збір репрезентативного та водночас повного бібліографічного показника у деякій області досліджень потребує тривалої роботи експертів.

### **Мета дослідження**

Поточна робота присвячена застосуванню запропонованого авторами методу побудови та аналізу мережі цитувань [11, 12] до укладання бібліографічного показника з теми "Методи персоналізації систем комп'ютерного навчання вимови".

### **Аналіз останніх досліджень і публікацій**

Наукові статті, які стосуються деякої області досліджень, можуть бути знайдені на основі ключових слів за допомогою запитів до баз даних, в збірниках праць спеціалізованих конференцій або журналів [7], на основі співавторства [4,5], на основі автоматичної класифікації текстів [6] та автоматично укладених онтологій [13], або методом снігової кулі [14, 15, 8, 9, 12].

Однак не для кожної теми дослідження існує відповідна конференція або журнал, онтології та автоматична класифікація текстів не показують наукової цінності праць, науковці можуть публікувати результати із кількох тем. Пошук за ключовими словами [16] страждає від індивідуальних варіацій тезаурусу. Кожен науковець – як автор наукових публікацій, так і експерт, який укладає бібліографічний показник, – використовує лише частину термінів своєї предметної області. І тому пошук на основі ключових слів знаходить не всі наукові публікації на задану тему.

Метод снігової кулі або "вибірка за ланцюжком знайомств" широко використовується в соціології [14, 17] для отримання репрезентативної вибірки із важкодоступних популяцій. Останнім часом метод снігової кулі знайшов використання в бібліографічних дослідженнях [15, 8, 9, 11, 12], де застосовується для формування бібліографічних показників. Аналогічно до оригінального методу снігової кулі, статті, які посилаються на початкові статті, формують перший рівень, потім статті, які посилаються на перший рівень, формують другий рівень і т.д. Ця техніка формує мережу релевантних статей, побудовану навколо зародку, яка надає набагато більше можливостей для аналізу, ніж простий упорядкований перелік – результат пошуку за ключовими словами. Дослідник контролює кількість рівнів, а також відбір публікацій для

наступного рівня снігової кулі. Було показано [18], що мережі цитування є безмасштабними, а тому у більшості випадків достатньо зібрати снігову кулю трьох рівнів [4].

Відбір публікацій для наступного рівня снігової кулі виконувався за допомогою Google PageRank [15, 8], косинусної міри схожості [9], імовірнісної статистичної моделі [11, 12]. Також можливим є застосування методів машинного навчання із вчителем [19]. Кандидатами на міру впливовості публікацій є традиційні показники: індекс цитування та індекс Гірша [20], кількість шляхів цитування [21].

Індекс Гірша для автора кількісно дорівнює числу його статей, кількість посилань на які у базах даних дорівнює або перевищує значення індексу Гірша. Індекс Гірша є кількісною характеристикою продуктивності вченого за весь період наукової діяльності, і тому він не може бути використаний для оцінки впливу однієї наукової публікації на розвиток окремої області знання.

Індекс цитування – прийнята в науковому світі міра значущості наукової роботи будь-якого науковця або наукового колективу [22]. Величина індексу цитування визначається кількістю посилань на публікацію, або на прізвище автора в інших джерелах. PageRank [23] – сімейство алгоритмів оцінки важливості веб-сторінок, які враховують як кількість посилань на документ, так і вагу документів, які посилаються на задану веб-сторінку. Звичайно PageRank та індекс цитування обчислюються для всіх проіндексованих документів, тому їх використання приводить до завищеної оцінки класичних робіт, довідників, підручників, які стосуються не обраної теми, а суміжних областей знань.

Перспективними для селекції є методи, що спираються на зміст публікацій. Основною ідеєю таких підходів є порівняння кандидатів на включення до снігової кулі із статтями, що складають зародок. Для порівняння використовуються підходи застосування косинусної міри схожості [9], різноманітні міри схожості рядків [24] та імовірнісні статистичні моделі [11, 12]. Перший підхід використовує векторну модель документа та порівнює множини слів у двох текстах, другий спосіб вимірює схожість написання, третій використовує один із різновидів тематичного моделювання [25]. Тематичні моделі дозволяють порівнювати зміст текстів, враховувати синоніми та виявляти зв'язки між термінами.

Ще одним способом відбору найважливіших документів із уже зібраної колекції публікацій є аналіз отриманої мережі цитування – напрямленого графа, у якому вузлами є публікації, а ребрами – посилання з однієї публікації на іншу. Аналіз основного шляху, вперше запропонований Гуммоном і Доріаном [21], частково автоматизований засобами програмного пакета Pajek [26] та повністю автоматизований одним із авторів статті [27], виявляє найбільш значимі ланцюжки цитувань. Спочатку для кожного посилання обчислюється кількість шляхів, які проходять через нього, а потім відбираються посилання із найбільшою кількістю шляхів. Метод застосовний до будь-якої діяльності людини, яка може бути представлена у вигляді напрямленого графа: для відстеження шляхів потоку знань або траєкторій розвитку галузі науки і техніки за допомогою бібліографічних цитат або патентних цитат; до судових рішень, щоб відстежити зміну юридичних висновків; під час укладання бібліографічних покажчиків [28, 29, 30, 31, 10, 15, 8].

Застосований у поточній роботі метод збору бібліографічного покажчика [12] поєднує в одній інформаційній технології метод снігової кулі, тематичне моделювання та аналіз мережі цитування методом підрахунку шляхів, що дозволяє автоматизувати процес складання бібліографічних покажчиків і отримати незміщений повний і стійкий перелік літератури на задану тему.

### **Викладення основного матеріалу досліджень**

#### **Метод побудови та аналізу мережі цитування**

Загальна послідовність методу обмеженої снігової кулі [32, 12] містить наступні кроки:

1. Збір початкової множини публікацій – зародку вибірки – і її додавання до черги нульового рівня.
2. Ітерації звичайного алгоритму снігової кулі з метою збору текстів для побудови тематичної моделі. Для  $n \in 0, 1, 2, 3$  виконуються наступні кроки:
  - 2.1 вибірка порції публікацій із  $n$ -ї черги;
  - 2.2 завантаження публікацій, які посилаються на вибрану порцію;
  - 2.3 завантаження публікацій, на які посилаються статті із вибраної порції;
  - 2.4 додавання всіх завантажених публікацій у  $(n + 1)$ -у чергу.
3. Створення тематичної моделі на основі завантажених на попередньому кроці публікацій:
  - 3.1 отримання заголовків та анотацій всіх документів;
  - 3.2 розділення всіх заголовків та анотацій на речення;
  - 3.3 створення словника термінів, який містить всі знайдені у реченнях іменники та прикметники;
  - 3.4 утворення із термінів, які зустрічаються в одному реченні, всіх можливих пар та підрахунок їх кількості з метою оцінки сумісної ймовірності;
  - 3.5 знаходження стоп-слів, характерних для предметної області, та виключення їх з подальшого розгляду;
  - 3.6 виконання розрідженої симетричної невід'ємної факторизації матриці сумісної ймовірності [33] для визначення параметрів тематичної моделі;
  - 3.7 відображення кожної із статей зародку вибірки у вектор ймовірностей тем.

4. Виконання пакетного завантаження методом обмеженої снігової кулі: для  $n \in 0, 1, 2, 3$ 
  - 4.1. вибірка порції публікацій із  $n$ -ї черги;
  - 4.2. завантаження публікацій, які посилаються на вибрану порцію;
  - 4.3. завантаження публікацій, на які посилаються статті із вибраної порції;
  - 4.4. перетворення кожного завантаженого документа на мішок слів та стемінг;
  - 4.5. відображення кожної із завантажених статей у вектор ймовірностей тем;
  - 4.6. обчислення міри відмінності між статтями із зародку вибірки та завантаженими статтями;
  - 4.7. додавання в  $(n + 1)$ -у чергу тих завантажених публікацій, які є достатньо близькими до зародку вибірки, у черзі зберігаються назва, анотація, посилання та ймовірності тем для кожної статті.
5. Аналіз отриманої мережі цитування
  - 5.1 визначення ваги кожного посилання та документа методом підрахунку шляхів пошуку;
  - 5.2 вибір найвагоміших глобальних шляхів у мережі цитування.

Результатом аналізу є скорочена мережа цитування, яка містить тільки найважливіші наукові публікації з обраної теми.

#### Укладання бібліографічного покажчика

Збір початкової множини публікацій - зародку вибірки - відбувався за допомогою пошуку за ключовими словами "PERSONALIZATION IN COMPUTER-AIDED PRONUNCIATION TRAINING" в базі наукових публікацій SemanticScholar. Із знайдених публікацій було обрано 20 статей, які здалися підходящими за тематикою, наприклад [3, 35, 36, 38, 53, 54, 60, 65]. Простий метод снігової кулі завантажив 65035 анотацій та заголовків статей, на основі яких було складено словник об'ємом 10453 термінів, із яких 7348 були позначені як рідкісні слова і 62 - як специфічні для зібраних текстів стоп-слова. У скорочений словник потрапили 3045 термінів, які містять 95% всіх слів у зібраній колекції. До рідкісних слів потрапили всі помилкові терміни та рідкісні словосполучення. Стоп-словами виявились характерні для предметної області терміни "operation", "major", "properties", "computation", "processing", "particular", "number", "addit", "details" і т.д. Пакетне завантаження методом обмеженої снігової кулі дозволило створити мережу цитування із 466 публікацій, у якій було знайдено 50 найголовніших шляхів цитування, які проходять через 43 статті, що і складають бібліографічний покажчик.

#### Аналіз бібліографічного покажчика

Було проведено широке дослідження дизартрії, яка зазвичай містить більшість дефектів вимови [1]. Системи автоматичного розпізнавання мови (САРМ), які спрямовані на "розуміння" дизартричної вимови можуть бути згруповані за наступними напрямками: корекція звукового сигналу, використання знань про артикуляцію, використання фонологічних особливостей мови, моделювання вимови, а також адаптація лексичних, акустичних та математичних моделей. Нами також були виявлені декілька суміжних областей досліджень, які пов'язані з адаптацією до дизартричної вимови.

Роботи, в яких застосовуються знання про артикуляцію були розглянуті в роботах [2, 34-38]. В [2] описується обчислювальна модель вимови людини, яка спроможна вловлювати ефекти від найменших акустичних відмінностей при розпізнаванні вимови. САРМ з акустичною моделлю, що базується на субфонетичних частинах і багатьох фонологічних особливостях була запропонована в [34]. Басова мережа, що застосовується для розпізнавання послідовностей фонем розглянута в [35]. Рекурентні нейронні мережі та нейронні мережі прямого поширення, а також радіально базовані та послідовно базовані опірні вектори для розпізнавання дизартричної вимови було запропоновано в [36]. Емпіричні знання з голосового тракту при дефектах вимови, таких як дизартрична вимова, були досліджені в [37]. Інверсія від акустичної складової до артикуляційної для оцінки положення голосового тракту, використовуючи нелінійну систему Хаммерштейна була запропонована в [38].

Про адаптацію математичних моделей було звітано в [39-49], починаючи з обчислення мінімуму залишку передбачення [39]. Приховані Марковські моделі (Hidden Markov Models, НММ) були використані для декомпозиції сигналів в [40]. Адаптація спікера, що базується на прихованих Марковських моделях неперервної щільності (СДНММ), де параметри моделі сформульовані як Басова навчальна процедура, досліджена в [41]. НММ для розпізнавання ізольованих слів у автоматичній моделі з великою кількістю варіацій вимови було розглянуто в [42]. У [43] була презентована стратегія адаптації, яка базується на кусково-лінійній функції, параметрами якої є характеристики спікера та характеристики зразкового спікера. Метод максимальної правдоподібності лінійної регресії адаптації спікера для СДНММ була запропонована в [44]. Проблема з кількістю даних для адаптації була вирішена в [45], за допомогою техніки стриманого оцінювання для моделі Гаусової суміші. Акустичне моделювання для САРМ за допомогою лінійної динамічної моделі, тобто підвиду моделі простору станів запропоновано в [46]. Адаптація до спікерів зі спастичною дизартрією шляхом використання ПММ та методу опорних векторів розглянута в [47]. Математичне моделювання також використовується для виправлення помилок в розпізнаванні дизартричної вимови, як показано у [48], де перетворювач вагових коефіцієнтів, що базується на контексті вимови слугує цій цілі. У [49] був наданий огляд декількох математичних підходів до адаптації розпізнавання вимови.

Методи адаптації звуку були розглянуті в [50-54]. Спочатку дослідження фокусувалися на виявленні та зменшенні фонового шуму [50]. В [51] було взято до уваги виявлення різномірних шумів. Метод представлення звукової форми мови в термінах абстрактних, лінгвістичних відмінностей для отримання множини дискримінаційних ознак для подальшого використання в мові було представлено в [52]. Перетворення звукової хвилі спікера з дефектами вимови в більш зрозумілий для слухачів вид був представлений у [53]. Використання вирівнювання для заходження довгих пауз та застосування інтервалів цих пауз, як опорних точок було запропоновано в [54].

Різноманітні акустичні методи, методи моделювання вимови, методи лексичної адаптації з використанням фонологічних особливостей та комбіновані методи були розглянуті нижче. Акустичне та лексичне моделювання дизартричної вимови було використано в [55]. Акустична адаптація разом зі знаннями про артикуляцію були використані для пристосування артикуляційної динаміки в розпізнаванні мови [3]. Матриця неточностей та словник звуків були використані в акустичній адаптації та артикуляційних знаннях у [56]. Принципи подолання проблеми підготовки бази зразків та архітектура системи розпізнавання вимови в системі з великим словниковим запасом були описані в [57]. Фонологічні властивості були розглянуті в [58], де була запропонована нейронна мережа для їх знаходження. У [59] було розглянуте розпізнавання мови шляхом фонетично розмічених складів для встановлення ефективності використання фонетичних особливостей у розпізнаванні фонем. Особлива увага була приділена базам зразків вимови, серед яких TORGUS [60], база дизартричної вимови Nemours [61], AVICAR – аудіовізуальна база зразків записаних в автомобілі [62] та DARPA – база однієї тисячі слів на тематику менеджменту [63]. На кінець в [64-65] був розглянутий метод підбору спікерів для найбільш правильного тренування акустичної моделі дизартричної вимови.

### Висновки

Робота показує практичне застосування запропонованої раніше інформаційної технології автоматичного створення бібліографічних покажчиків до теми “Методи персоналізації систем комп’ютерного навчання вимови”. Перевагою методу є автоматизація процесу складання переліку публікацій: отримуючи пошуковий запит у вигляді вибраних вручну зразків публікацій, він створює скорочену мережу цитування, яка містить найважливіші наукові публікації з обраної теми. Основними компонентами технології є ймовірнісне тематичне моделювання, вдосконалений метод снігової кулі та аналіз отриманої мережі цитування.

В результаті отримано та проаналізовано перелік публікацій з методів персоналізації систем комп’ютерного навчання вимови. Для адаптації застосовуються знання про артикуляцію, можливі варіації фонем та складів сама адаптація виконується або шляхом зміни математичних моделей - переважно прихованих Марківських моделей, Баєсівських або нейронних мереж, або трансформацією вхідного сигналу. Особлива увага була приділена базам зразків вимови. Майже всі методи адаптації перевірялися на дизартричній вимові - як на крайньому випадку порушення усного мовлення.

### Список використаної літератури

1. Gifford M. F. Speech defects and disorders // California and western medicine. — 1926. — Vol. 24, №. 1. — P. 72.
2. Scharenborg O. Towards capturing fine phonetic variation in speech using articulatory features / O. Scharenborg, V. Wan, R.K. Moore // Speech Communication. — 2007. — Vol. 49, №. 10-11. — P. 811-826.
3. Improving speech recognition for children using acoustic adaptation and pronunciation modeling. / Prashanth Gurunath Shivakumar, Alexandros Potamianos, Sungbok Lee, Shrikanth Narayanan // WOCCI. — 2014. — P. 15-19.
4. Newman M. E. The structure of scientific collaboration networks / M. E. Newman // Proceedings of the national academy of sciences. — 2001. — Vol. 98, №. 2. — P. 404-409.
5. Newman M. E. Coauthorship networks and patterns of scientific collaboration / M. E. Newman // Proceedings of the national academy of sciences. — 2004. — Vol. 101, №. suppl 1. — P. 5200-5205.
6. A new technique for building maps of large scientific domains based on the cocitation of classes and categories / Félix Moya-Anegón, Benjamín Vargas-Quesada, Victor Herrero-Solana et al. // Scientometrics. — 2004. — Vol. 61, №. 1. — P. 129-145.
7. Ontologies of time: Review and trends. / Vadim Ermolayev, Sotiris Batsakis, Natalya Keberle et al. // International Journal of Computer Science & Applications. — 2014. — Vol. 11, №. 3. — P. 57-115.
8. Mapping the historical development of physical activity and health research: A structured literature review and citation network analysis / Andrea Ramirez Varela, Michael Pratt, Jenine Harris et al. // Preventive medicine. — 2018. — Vol. 111. — P. 466-472.
9. Ahad A. Navigation through citation network based on content similarity using cosine similarity algorithm / A. Ahad, M. Fayaz, A. S. Shah // Int. J. Database Theory Appl. — 2016. — Vol. 9, №. 5. — P. 9-20.

10. Data envelopment analysis 1978–2010: A citation-based literature survey / John S Liu, Louis YY Lu, Wen-Min Lu, Bruce JY Lin // *Omega*. — 2013. — Vol. 41, №. 1. — P. 3–15.
11. Érdi P. Patent citation network analysis: Topology and evolution of patent citation networks /P. Érdi // *Artificial Neural Networks and Machine Learning – ICANN 2016, Proceedings of the 25th International Conference on Artificial Neural Networks, Lecture Notes in Computer Science, Vol. 9886*. —Springer. — 2016. — P.543.
12. Dobrovolskyi H. Collecting the seminal scientific abstracts with topic modelling, snowball sampling and citation analysis / H. Dobrovolskyi, N. Keberle // *CEUR Workshop Proceedings*. — 2018. — Vol. 2015. — P. 179-192.
13. Osborne F. Klink-2: integrating multiple web sources to generate semantic topic networks / F. Osborne, E. Motta // *The Semantic Web - ISWC 2015 (Part 1)*. — Springer. — 2015. — Vol. 9366. — P. 408–424.
14. Biernacki P. Snowball sampling: Problems and techniques of chain referral sampling / P. Biernacki, D. Waldorf // *Sociological methods & research*. — 1981. — Vol. 10, №. 2. — P. 141–163.
15. Leczy J. Representative literature reviews using constrained snowball sampling and citation network analysis [Електронний документ] / J. Leczy, K. Beatty // SSRN eLibrary. — 2012. — Режим доступу до ресурсу : <https://ssrn.com/abstract=1992601>.
16. Petticrew M. Planning and conducting systematic reviews /M. Petticrew, S. Gilbody // *Health Psychology in Practice*. — 2009. — P. 150–179.
17. Heckathorn D. D. Comment: Snowball versus respondent-driven sampling // *Sociological methodology*. — 2011. — Vol. 41, №. 1. — P. 355–366.
18. de Solla Price D. J. Networks of scientific papers // *Science*. — 1965. — P. 510–515.
19. Valenzuela M. Identifying meaningful citations / M. Valenzuela, V. Ha, O. Etzioni // *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*. — 2015.
20. Hirsch J. E. An index to quantify an individual’s scientific research output / J. E. Hirsch // *Proceedings of the National academy of Sciences*. — 2005. — Vol. 102, №. 46. — P. 16569–16572.
21. Hummon N. P. Connectivity in a citation network: The development of dna theory / N. P. Hummon, P. Dereian // *Social networks*. — 1989. — Vol. 11, №. 1. — P. 39–63.
22. Сухий О. Л. Методичні рекомендації “Алгоритми пошуку в інформаційних системах” [Електронний ресурс] / О.Л. Сухий, В.М.Міленін, В.М.Тарадайнік. // Інститут обдарованої дитини НАПН України — Режим доступу до ресурсу : [http://biblos.iod.gov.ua/docview.php?doc\\_id=2613](http://biblos.iod.gov.ua/docview.php?doc_id=2613).
23. The PageRank citation ranking: Bringing order to the web (Technical Report) / L. Page, S. Brin, R. Motwani, T. Winograd // *Stanford InfoLab*. — 1999.
24. Concordance Between the Findings of Epidemiological Studies and Randomized Trials in Nutrition: An Empirical Evaluation and Citation Analysis. Appendix B: Details on the construction of citation graphs / Denish Moorthy, Mei Chung, Jounghye Lee et al. // *Technical Review, №.17*. — Vol.6. — (Prepared by the Tufts Medical Center Evidence-based Practice Center under Contract №. 290-2007-10055-I.) AHRQ Publication №.13-EHC067-EF. Rockville, MD: Agency for Healthcare Research and Quality, May 2013.
25. Vorontsov K. Tutorial on probabilistic topic modeling: Additive regularization for stochastic matrix factorization / K. Vorontsov, A. Potapenko // *Communications in computer and information science*. — 2014. — Vol. 436. — P. 29–46.
26. Batagelj V. Efficient algorithms for citation network analysis. [Електронний ресурс] / V. Batagelj // *Cornell University Library* — 2003. — Режим доступу до ресурсу : <https://arxiv.org/abs/cs/0309023>.
27. Кучерян К. М. Реалізація алгоритму підрахунку шляхів пошуку для аналізу мережі цитування / К.М. Кучерян, Г.А. Добровольський // *Збірка тез доповідей Дев’ятої Всеукраїнської, шістнадцятої регіональної наукової конференції молодих дослідників «Актуальні проблеми математики та інформатики» Запоріжжя / ЗНУ*. — 2018. — P. 46.
28. Knowledge flows - Analyzing the core literature of innovation, entrepreneurship and science and technology studies / Samyukta Bhupatiraju, Önder Nomaler, Giorgio Triulzi, Bart Verspagen // *Research Policy*. — 2012. — Vol. 41, №. 7. — P. 1205–1218.
29. Calero-Medina C. Combining mapping and citation network analysis for a better understanding of the scientific development: The case of the absorptive capacity field / C.Calero-Medina, E.C. Noyons // *Journal of Informetrics*. — 2008. — Vol. 2, №. 4. — P. 272–279.
30. Colicchia C. Supply chain risk management: a new methodology for a systematic literature review / C. Colicchia, F. Strozzi // *Supply Chain Management*. — 2012. — Vol. 17, №. 4. — P. 403–418.
31. Mapping the multidisciplinary field of public health services and systems research / Jenine K. Harris, Kate E. Beatty, Jesse D. Leczy et al. // *American Journal of Preventive Medicine*. — 2011. — Vol. 41, №. 1. — P. 105–111.
32. Dobrovolskyi H. Probabilistic Topic Modelling for Controlled Snowball Sampling in Citation Network Collection / H. Dobrovolskyi, N. Keberle, O. Todoriko // *Communications in Computer and Information Science*. — 2017. — Vol. 786. — pp. 85-100.

33. Dobrovolskyi H. Sparse symmetric nonnegative matrix factorization applied to face recognition / H. Dobrovolskyi, N. Keberle, Y. Ternovyy // 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS). — 2017. — P. 1042–1045.
34. Metze F. Discriminative speaker adaptation using articulatory features / F. Metze // Speech Communication. — 2007. — Vol. 49, №. 5. — P. 348–360.
35. Rudzicz F. Applying discretized articulatory knowledge to dysarthric speech / F. Rudzicz // 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. — 2009. — P. 4501–4504.
36. Rudzicz F. Phonological features in discriminative classification of dysarthric speech / F. Rudzicz // Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on / IEEE. — 2009. — P. 4605–4608.
37. Rudzicz F. Articulatory knowledge in the recognition of dysarthric speech / F. Rudzicz // IEEE Transactions on Audio, Speech, and Language Processing. — 2011. — Vol. 19, №. 4. — P. 947–960.
38. Rudzicz F. Using articulatory likelihoods in the recognition of dysarthric speech / F. Rudzicz // Speech Communication. — 2012. — Vol. 54, №. 3. — P. 430–444.
39. Itakura F. Minimum prediction residual principle applied to speech recognition / F. Itakura // IEEE Transactions on Acoustics, Speech, and Signal Processing. — 1975. — Vol. 23, №. 1. — P. 67–72.
40. Varga A., Moore R. Hidden markov model decomposition of speech and noise / A. Varga, R. Moore // International Conference on Acoustics, Speech, and Signal Processing, ICASSP-90. — IEEE. — 1990. — P. 845–848.
41. Lee C.-H. A study on speaker adaptation of the parameters of continuous density hidden markov models / C.-H. Lee, C.-H. Lin, B.-H. Juang // IEEE Transactions on Signal Processing. — 1991. — Vol. 39, №. 4. — P. 806–814.
42. Deller Jr J. On the use of hidden markov modelling for recognition of dysarthric speech / J. Deller Jr., D. Hsu, L. J. Ferrier // Computer Methods and Programs in Biomedicine. — 1991. — Vol. 35, №. 2. — P. 125–139.
43. Robust speaker adaptation using a piecewise linear acoustic mapping / Jerome R Bellegarda, Peter V de Souza, Arthur J Nadas et al. // International Conference on Acoustics, Speech, and Signal Processing, ICASSP-92. — IEEE. — Vol. 1. — 1992. — P. 445–448.
44. Leggetter C. J. Maximum likelihood linear regression for speaker adaptation of continuous density hidden markov models / C. J. Leggetter, P. C. Woodland // Computer speech & language. — 1995. — Vol. 9, №. 2. — P. 171–185.
45. Digalakis V. V. Speaker adaptation using constrained estimation of gaussian mixtures / V. V. Digalakis, D. Rtischev, L. G. Neumeyer // IEEE Transactions on speech and Audio Processing. — 1995. — Vol. 3, №. 5. — P. 357–366.
46. Frankel J. Linear dynamic models for automatic speech recognition : PhD thesis / Frankel J. — University of Edinburgh — 2004. — 355 p.
47. Hmm-based and svm-based recognition of the speech of talkers with spastic dysarthria / Mark Hasegawa-Johnson, Jon Gunderson, Adrienne Perlman, Thomas Huang // International Conference on Acoustics, Speech and Signal Processing, ICASSP 2006. — IEEE. — Vol. 3. — 2006. — P. III–III.
48. Seong W. K., Park J. H., Kim H. K. Dysarthric speech recognition error correction using weighted finite state transducers based on context-dependent pronunciation variation // Lecture Notes in Computer Science. — 2012. — Vol. 7383. — P. 475–482.
49. Shinoda K. Speaker adaptation techniques for automatic speech recognition / K. Shinoda // Proceedings APSIPA ASC. — 2011.
50. A. Speech recognition using noise-adaptive prototypes / A. Nádas, D. Nahamoo, M. A. Picheny // IEEE Transactions on Acoustics, Speech, and Signal Processing. — 1989. — Vol. 37, №. 10. — P. 1495–1503.
51. Hermansky H. Rasta processing of speech / H. Hermansky H., N. Morgan // IEEE transactions on speech and audio processing. — 1994. — Vol. 2, №. 4. — P. 578–589.
52. Eide E. Distinctive features for use in an automatic speech recognition system / E. Eide // Seventh European Conference on Speech Communication and Technology. — 2001.
53. Rudzicz F. Adjusting dysarthric speech signals to be more intelligible / F. Rudzicz // Computer Speech & Language. — 2013. — Vol. 27, №. 6. — P. 1163–1177.
54. Yeung Y.T. Improving automatic forced alignment for dysarthric speech transcription / Y.T. Yeung, K.H. Wong, H. Meng // Sixteenth Annual Conference of the International Speech Communication Association. — 2015.
55. Mengistu K.T. Adapting acoustic and lexical models to dysarthric speech / K.T. Mengistu, F. Rudzicz // 2011 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP-2011. — IEEE. — 2011. — P. 4924–4927.

56. Rudzicz F. Correcting errors in speech recognition with articulatory dynamics / F. Rudzicz // Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics / Association for Computational Linguistics. — 2010. — P. 60–68.
57. Young S. A review of large-vocabulary continuous-speech / S. Young // IEEE signal processing magazine. — 1996. — Vol. 13, №. 5. — P. 45.
58. King S. Detection of phonological features in continuous speech using neural networks / S. King, P. Taylor // Computer Speech & Language. — 2000. — Vol. 14, №. 4. — P. 333–353.
59. Speech recognition via phonetically-featured syllables / Simon King, Paul Taylor, Joe Frankel, Korin Richmond. — 2000.
60. Rudzicz F. The TORGO database of acoustic and articulatory speech from speakers with dysarthria /F. Rudzicz, A. K. Namasivayam, T. Wolff // Language Resources and Evaluation. — 2012. — Vol. 46, №. 4. — P. 523–541.
61. The Nemours database of dysarthric speech / X. Menéndez-Pidal, J.B. Polikoff, S.M. Peters, J.E. Leonzio, H.T. Bunnell et al. // Proceedings of the Fourth International Conference on Spoken Language Processing. IEEE. — 1996.
62. Avicar: Audio-visual speech corpus in a car environment / Bowon Lee, Mark Hasegawa-Johnson, Camille Goudeseune et al. // Eighth International Conference on Spoken Language Processing. — 2004.
63. The DARPA 1000-word resource management database for continuous speech recognition / Patti Price, William M Fisher, Jared Bernstein, David S Pallett // Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on / IEEE. — 1988. — P. 651–654.
64. Lippmann R. P. Speech recognition by machines and humans // Speech communication. — 1997. — Vol. 22, №. 1. — P. 1–15.
65. Mengistu K. T. Comparing humans and automatic speech recognition systems in recognizing dysarthric speech /K.T. Mengistu, F. Rudzicz // Canadian Conference on Artificial Intelligence / Springer. — 2011. — P. 291–300.