

Запорожский национальный технический университет
Кафедра системного анализа и вычислительной математики

Латентно-семантический анализ в наукометрических сетях

Докладчик:

Пасечник М.С., аспирант

Научный руководитель:

Бакурова А.В.,
доктор экон. наук,
профессор кафедры СА и ВМ

Латентно-семантический анализ в наукометрических сетях

Актуальность:

в связи с ростом количества текстовой информации, необходим анализ этой информации. Использование информационных систем для решения этой задачи, делает актуальной разработку математических методов анализа текстовой информации для принятия решений в управлении социально-экономическими системами.

Цель работы:

проанализировать опыт стран СНГ, используя систему Google Scholar, изучив математические методы используемые для анализа текстовой информации, оценить популярность методов, сравнить с популярностью метода латентно-семантического анализа зарубежом.

Латентно-семантический анализ в наукометрических сетях

Латентно-семантический анализ (ЛСА) (англ. Latent semantic analysis, LSA) — это метод обработки информации на естественном языке, анализирующий взаимосвязь между коллекцией документов и терминами в них встречающимися, сопоставляющий некоторые факторы (тематики) всем документам и терминам.

В основе метода латентно-семантического анализа лежат принципы факторного анализа, в частности, выявление латентных связей изучаемых явлений или объектов. При классификации / кластеризации документов этот метод используется для извлечения контекстно-зависимых значений лексических единиц при помощи статистической обработки больших объемов текстов.

Латентно-семантический анализ в наукометрических сетях

Простейшая модель текста "**мешок слов**" (**bag-of-words**) представляет собой суммативное единство (не систему) составляющих текст слов.

Основной объект модели мешка слов — это слово, снабженное единственным атрибутом, частотой встречаемости этого слова.

В модели текста "мешок слов" учитывается только количество вхождений конкретных слов в исходном тексте, при этом игнорируются:

- порядок слов в документе,
- морфологические формы представления слов.

Латентно-семантический анализ в наукометрических сетях

Поиск статей проводился в базе Google Scholar по ключевым словосочетаниям:

- латентно-семантический анализ;
- мешок слов.

Наукометрическая база Google Scholar используется на постсоветском пространстве для общей оценки результативности научной деятельности как отдельных сотрудников, так и кафедр, факультетов в целом на основе индекса Хирша — h -index

Латентно-семантический анализ в наукометрических сетях

Таблица 1 - Статьи со словами «латентно-семантический анализ»

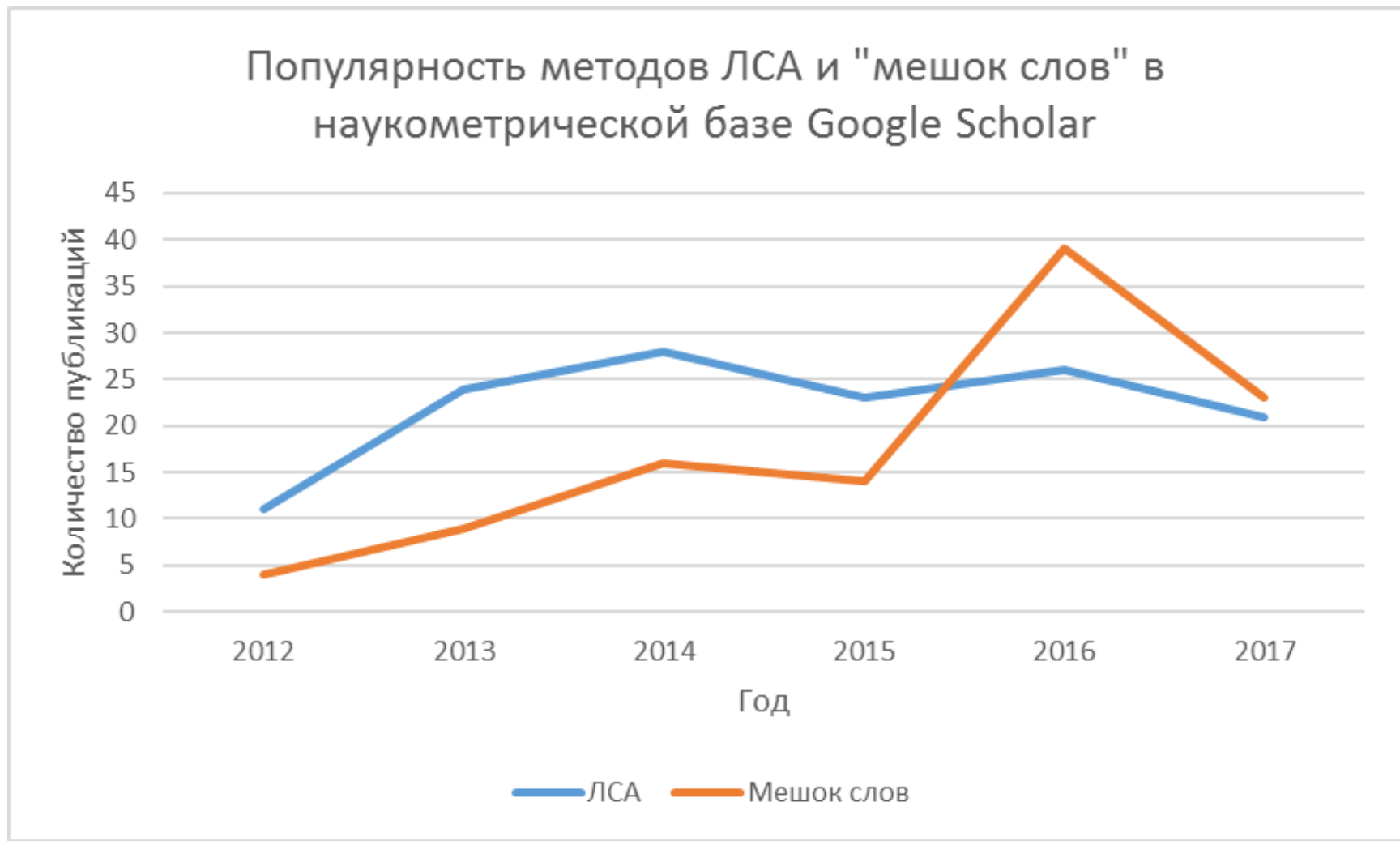
Год	Количество статей	География издательства
2012	11	1 — Киргизия 3 — Украина, 7 — Россия
2013	24	1 — Беларусь 6 — Украина, 17 - Россия
2014	28	6 — Украина, 22 - Россия
2015	23	4 — Украина, 19 — Россия
2016	26	3 — Украина, 23 — Россия
2017	21	3 — Украина, 18 — Россия

Латентно-семантический анализ в наукометрических сетях

Таблица 2 - Статьи со словами «мешок слов»

Год	Количество статей	География издательства
2012	4	4 — Россия
2013	9	7 — Россия 2 — Украина
2014	16	16 — Россия
2015	14	13 — Россия 1 — Украина
2016	39	39 — Россия
2017	23	23 — Россия

Латентно-семантический анализ в наукометрических сетях



Латентно-семантический анализ в наукометрических сетях

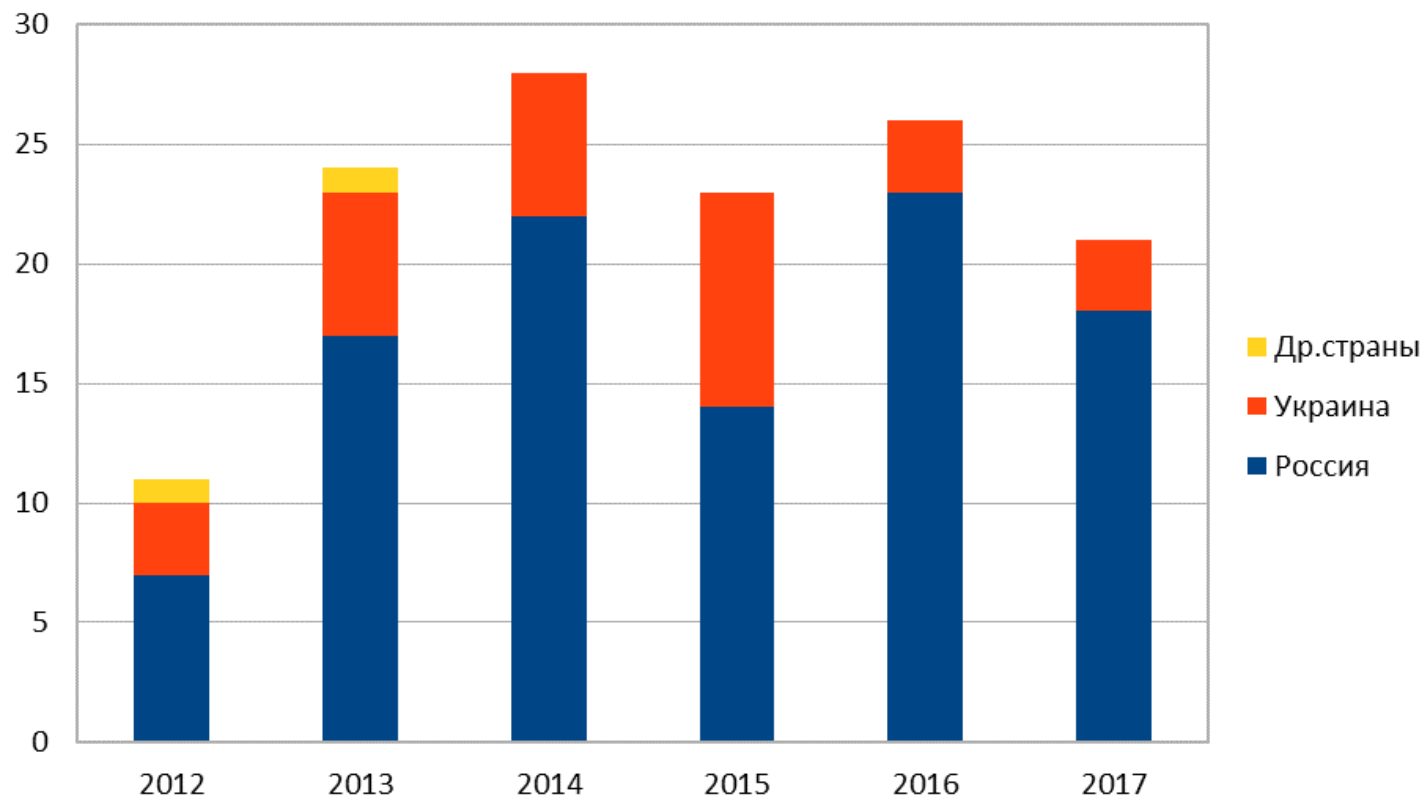


Рисунок 2 - Использование термина «латентно-семантический анализ» в различных странах, издающих журнал

Латентно-семантический анализ в наукометрических сетях

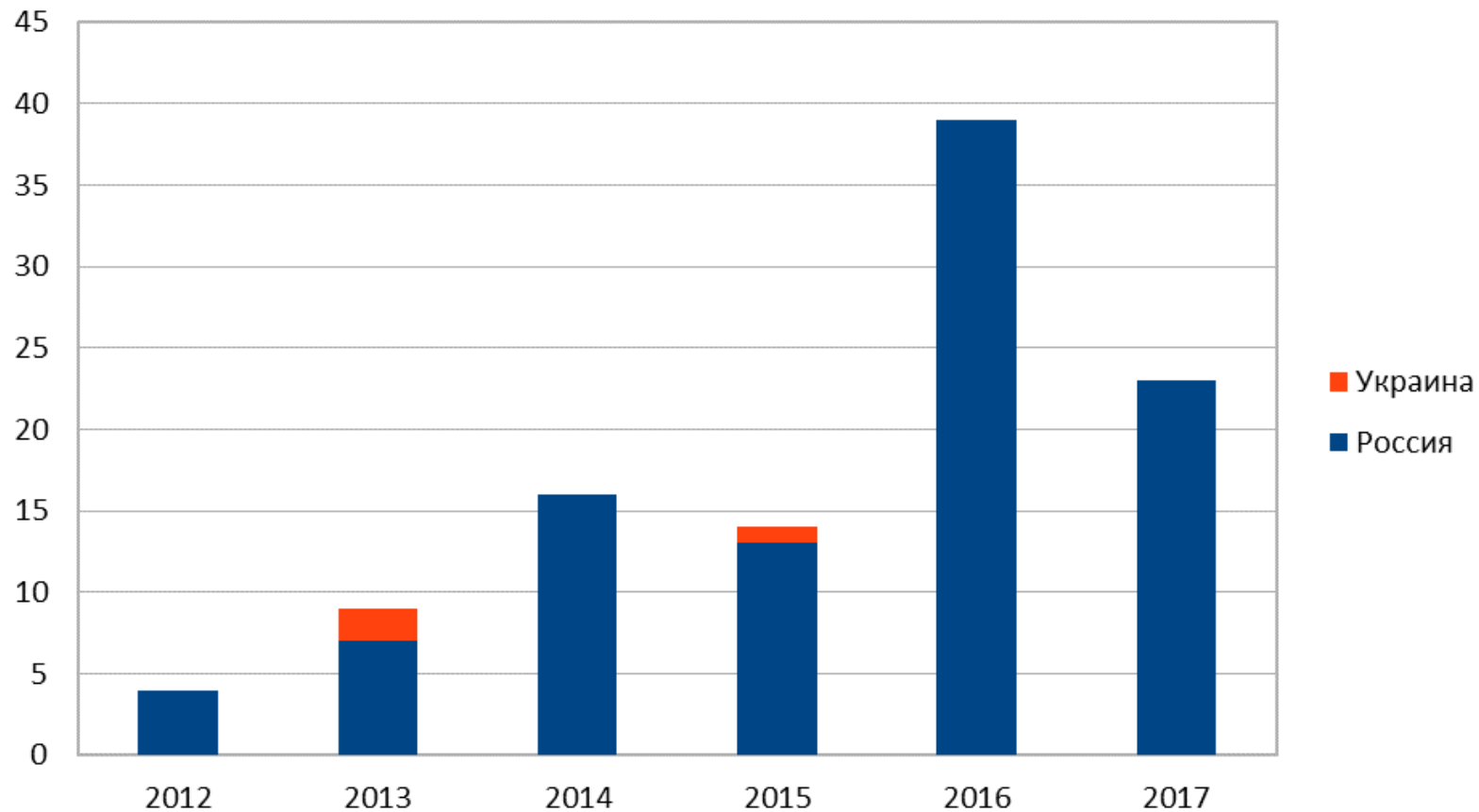


Рисунок 3 - Использование термина «мешок слов» в различных странах, издающих журнал

Латентно-семантический анализ в наукометрических сетях

Выводы

Методы латентно-семантического анализа и «мешок слов», как исследовательский инструмент, используются учеными во многих областях науки. Исследования наукометрической сети Google Scholar для поиска информации об использовании латентно-семантического анализа и метода «мешок слов» показали, что наибольшее число статей с использованием этих терминов на постсоветском пространстве было издано в России, в меньшей степени были представлены издательства Украины.

В последние годы виден значительный рост количества статей с этими терминами, а значит, распространение ЛСА продолжается, и ЛСА становится неотъемлемой частью научного мира.