

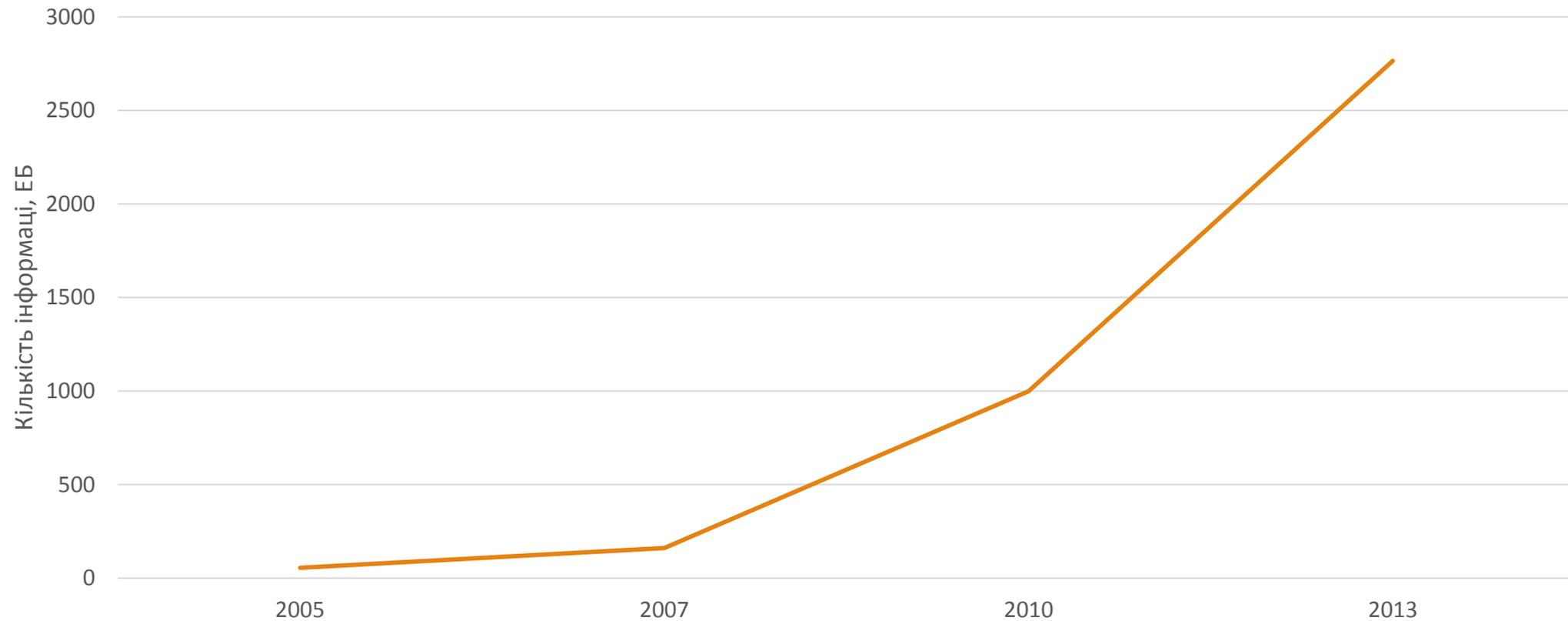
Структурирование знаний с использованием технологий интеллектуального анализа текста

КАЛЮЖНЫЙ МАКСИМ ДА-21

Цели работы

- Рассмотреть возможные этапы интеллектуального анализа текста.
- Рассмотреть теоретически алгоритмы для каждого из этапов.
- Разработать алгоритмы для каждого из этапов.
- Провести анализ эффективности работы алгоритмов

Количество информации, созданное человечеством



Требования к системе

- Точность
- Эффективность
- Скорость разработки
- Гибкость
- Устойчивость
- Масштабируемость
- Разреженность данных
- Мультиязычность

Основные этапы интеллектуального анализа текстов



Предварительная обработка

- Удаление стоп слов
- Стемминг
- N-граммы
- Приведение регистра, удаление пунктуации

Классификация методов поиска ключевых слов

Статистические

TF-IDF

Частотный

F-measure

Лингвистические

Графовые

Словарные

Маркmemный анализ

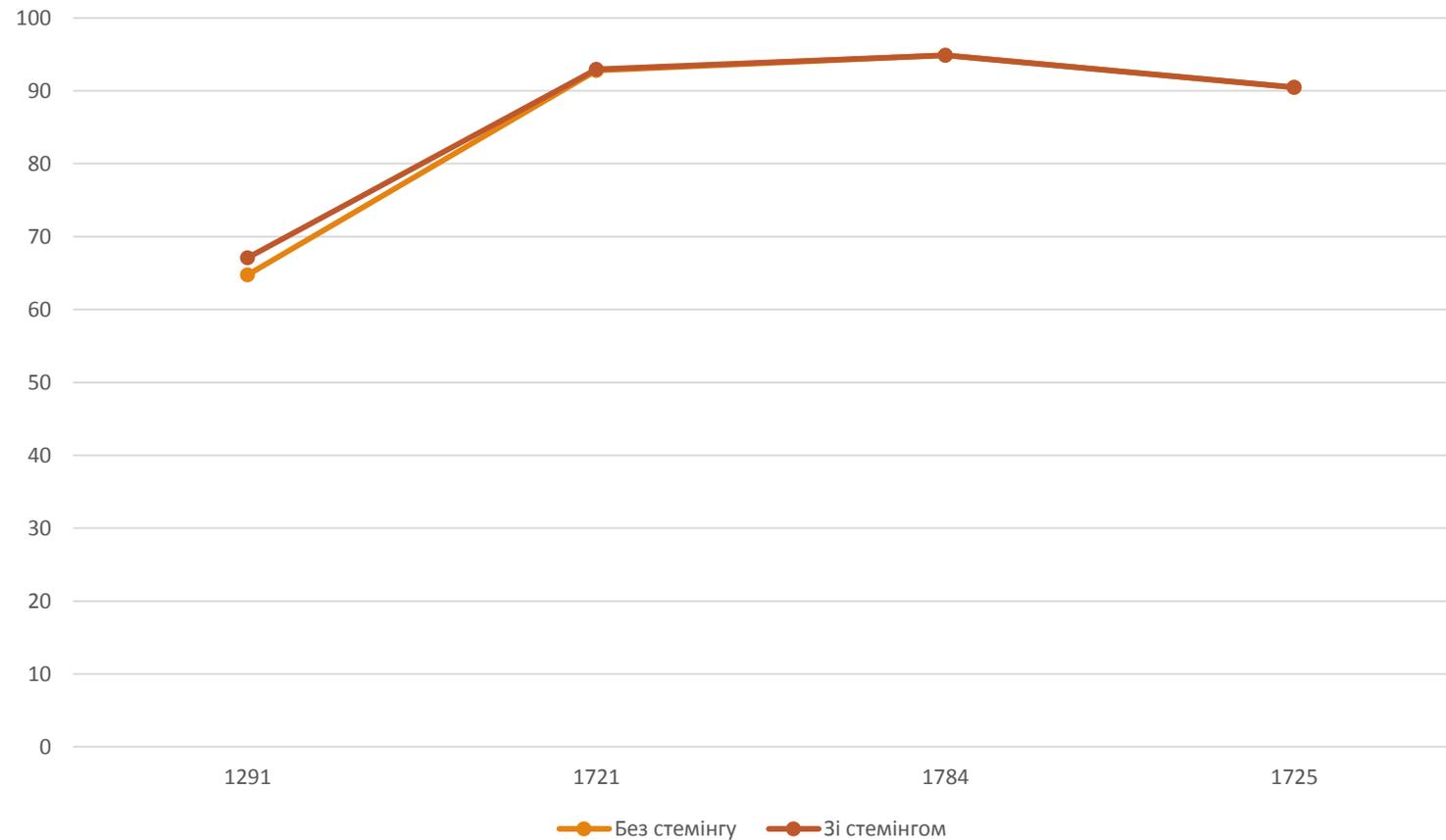
Гибридные

Лингво-статистических
шаблонов

Удаление стоп-слов

- Статистический
- Словарный
- Построенный на γ -интерпретации закона Бредфорда с предварительным стемингом
- Построенный на γ -интерпретации закона Бредфорда без предварительного стеминга

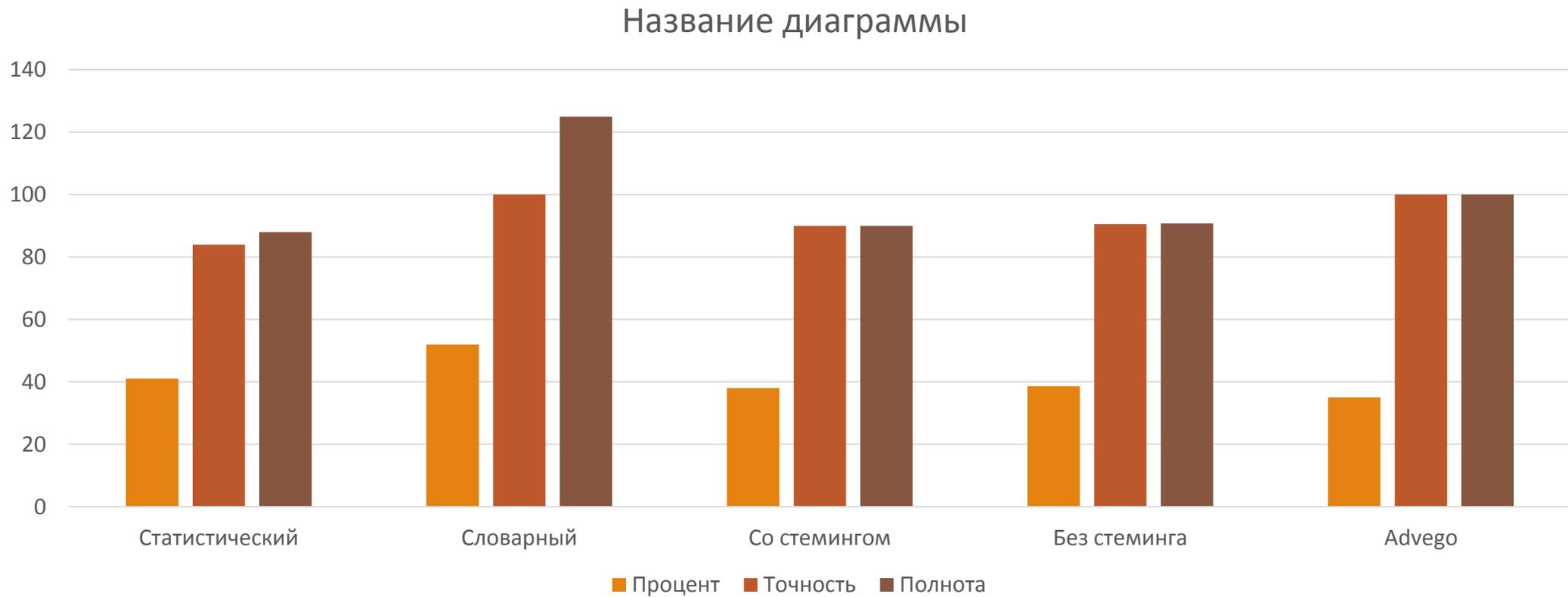
Зависимость между точностью и количеством слов для γ -интерпретации закона Бредфорда



Зависимость между полнотой и количеством слов для Y-интерпретации закона Бредфорда



Эффективность методов удаления СТОП-СЛОВ



Mepa TF-IFD



F-measure



Метод лингво-статистических шаблонов

Спасибо за внимание