КОРПОРАЦИЯ Галактика®

“Галактика-Zoom” на РОМИП’2006

А.В. Антонов,
С.Г. Баглей,
В.С. Мешков

{alexa, baglei, meshkov}
@galaktika.ru

Участие в дорожках РОМИП’2006

- Классификация нормативно-правовых документов
- Классификация Веб-страниц
- Классификация Веб-сайтов
- Поиск документов по документу-образцу

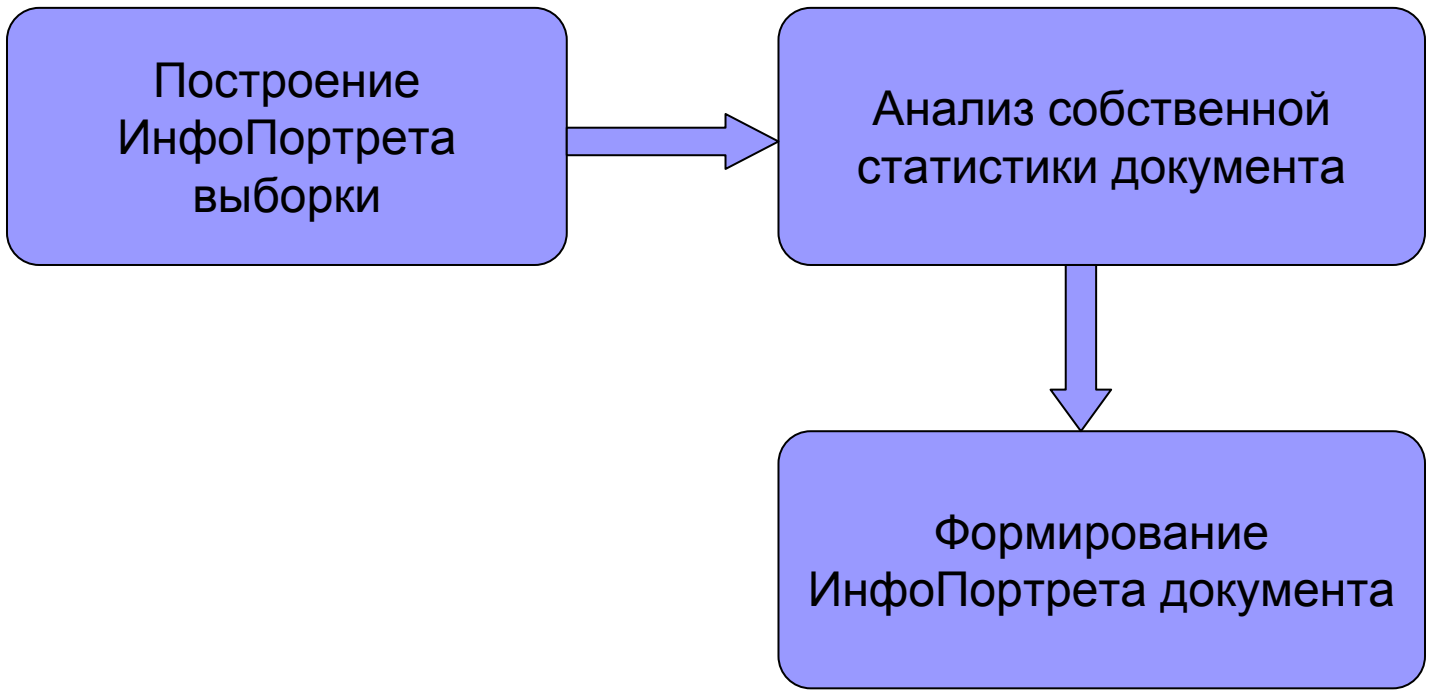
Технологии для классификации и поиска

- Информационный Портрет (ИнфоПортрет) выборки документов
- Ранжирование документов на основе ИнфоПортрета
- Классификация с использованием метода опорных векторов
- Кластеризация с использованием модифицированного метода латентного семантического анализа

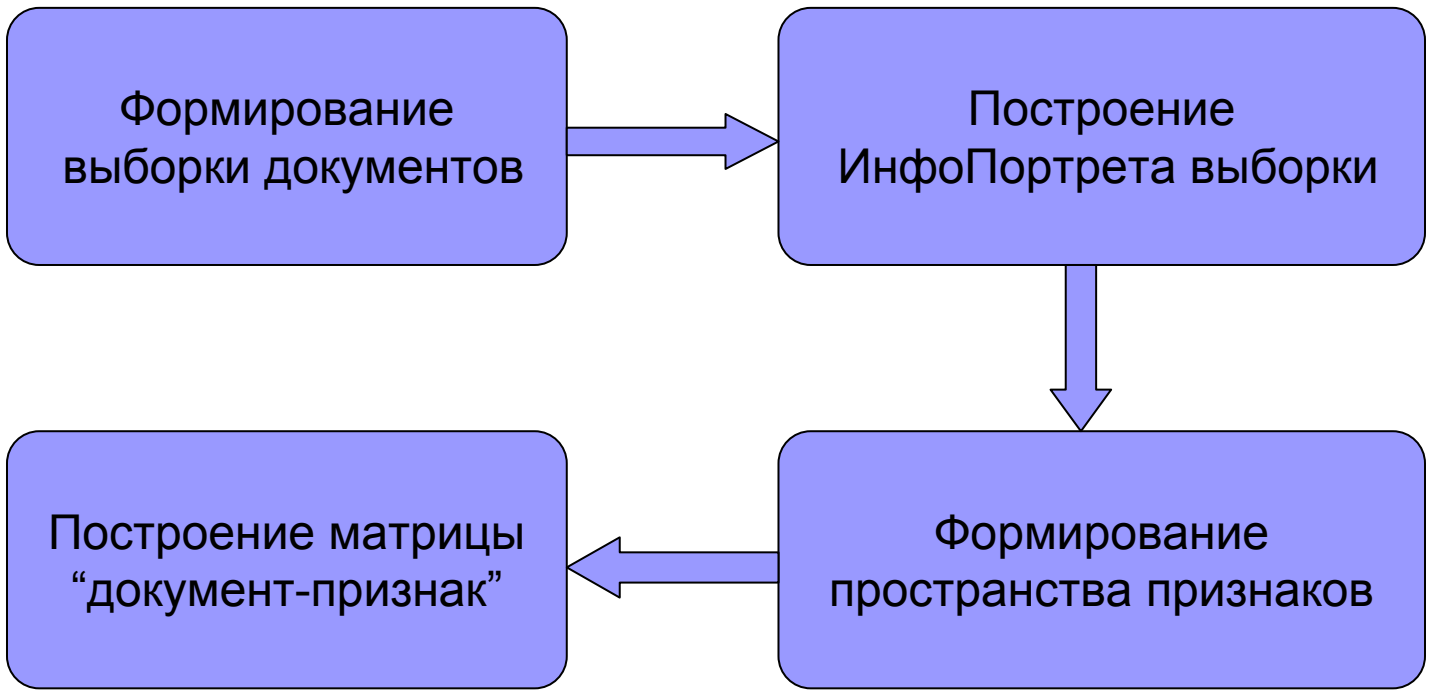
ИнфоПортрет выборки документов по запросу: “Аортокоронарное шунтирование”

СОСУД	АРТЕРИЯ
БОЛЬНОЙ	ХИРУРГ
ИШЕМИЧЕСКАЯ БОЛЕЗНЬ	КЛИНИЧЕСКИЙ
ИНФАРКТ	БОЛЬНАЯ
КОРОНАРНЫЙ	АТЕРОКЛЕФИТ
ХИРУРГИЯ	КАРДИОЛОГИЯ
ДОЛЕЧИВАНИЕ	СОЦИАЛЬНОЕ СТРАХОВАНИЕ
МИОКАРД	КЛИНИКА
САНАТОРИЙ	КАРДИОХИРУРГ
ИШЕМИЧЕСКИЙ	КРОВООБРАЩЕНИЕ
ОСТРЫЙ ИНФАРКТ	СПЕЦИАЛИЗИРОВАННЫЕ САНАТОРИИ
КОРОНАРНЫЕ АРТЕРИИ	КОРОНАРНЫЕ СОСУДЫ

Отображение документа



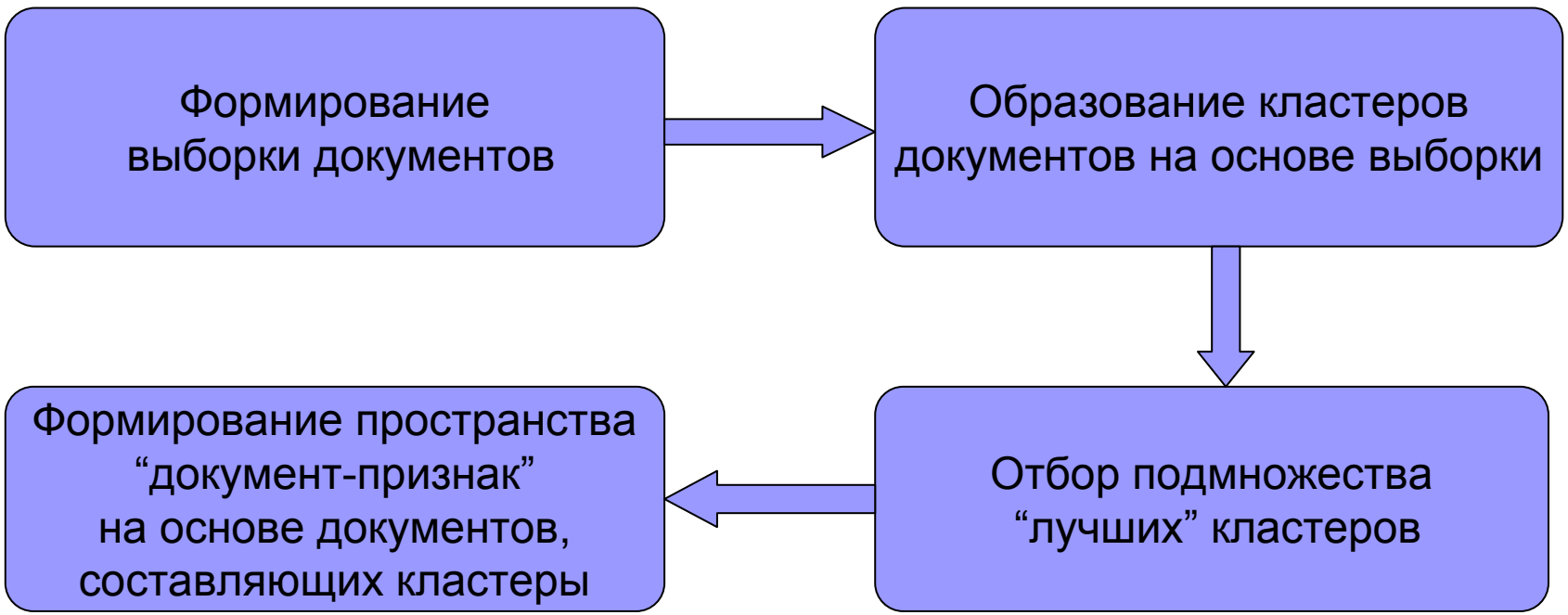
Отображение множества документов



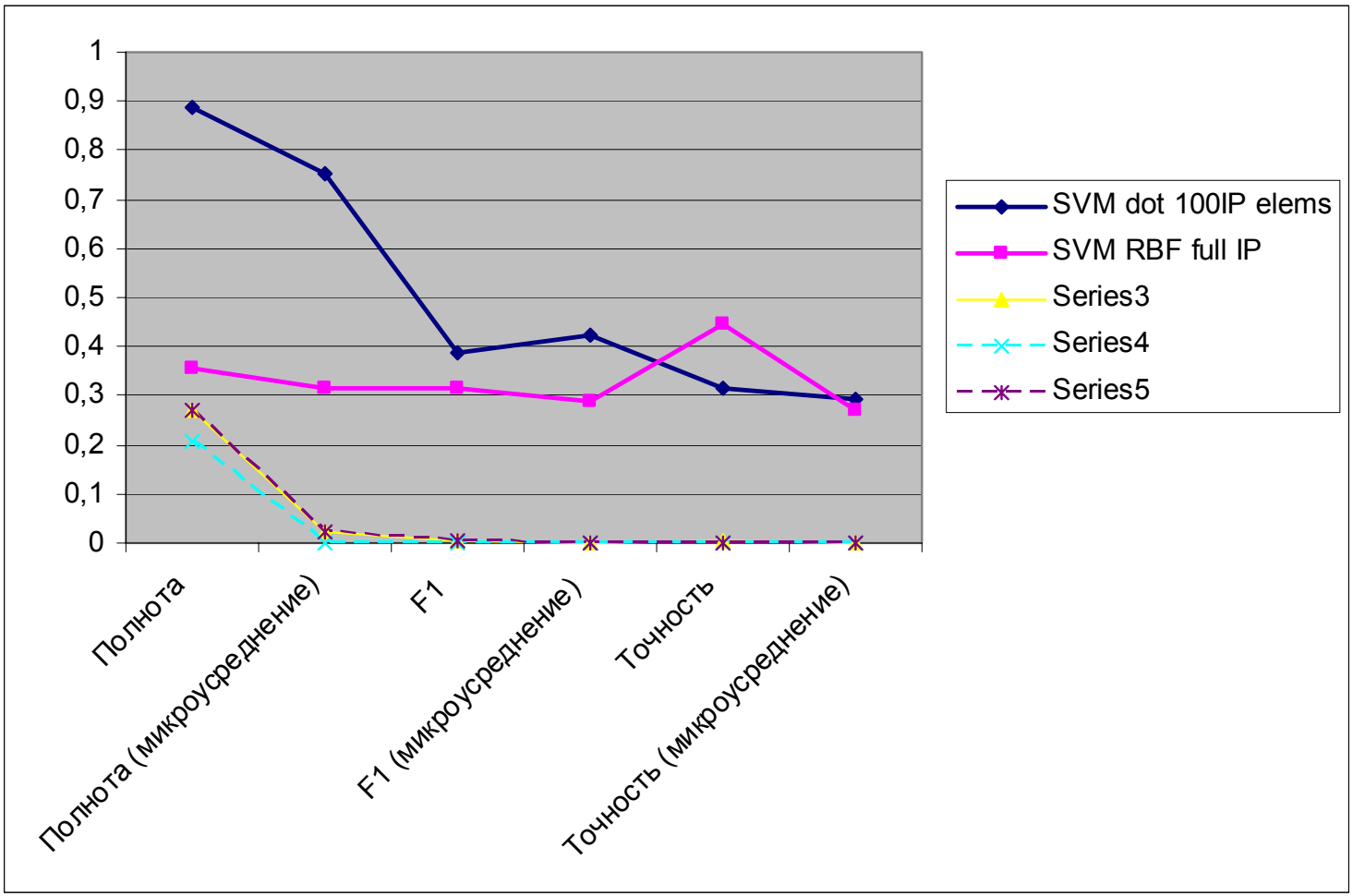
SVM

- Использована реализация метода SVMLight (T. Joachims, 1998)
- Для обработки данных выбраны линейное и гауссово ядро
- Прямой и регрессивный режим работы алгоритма

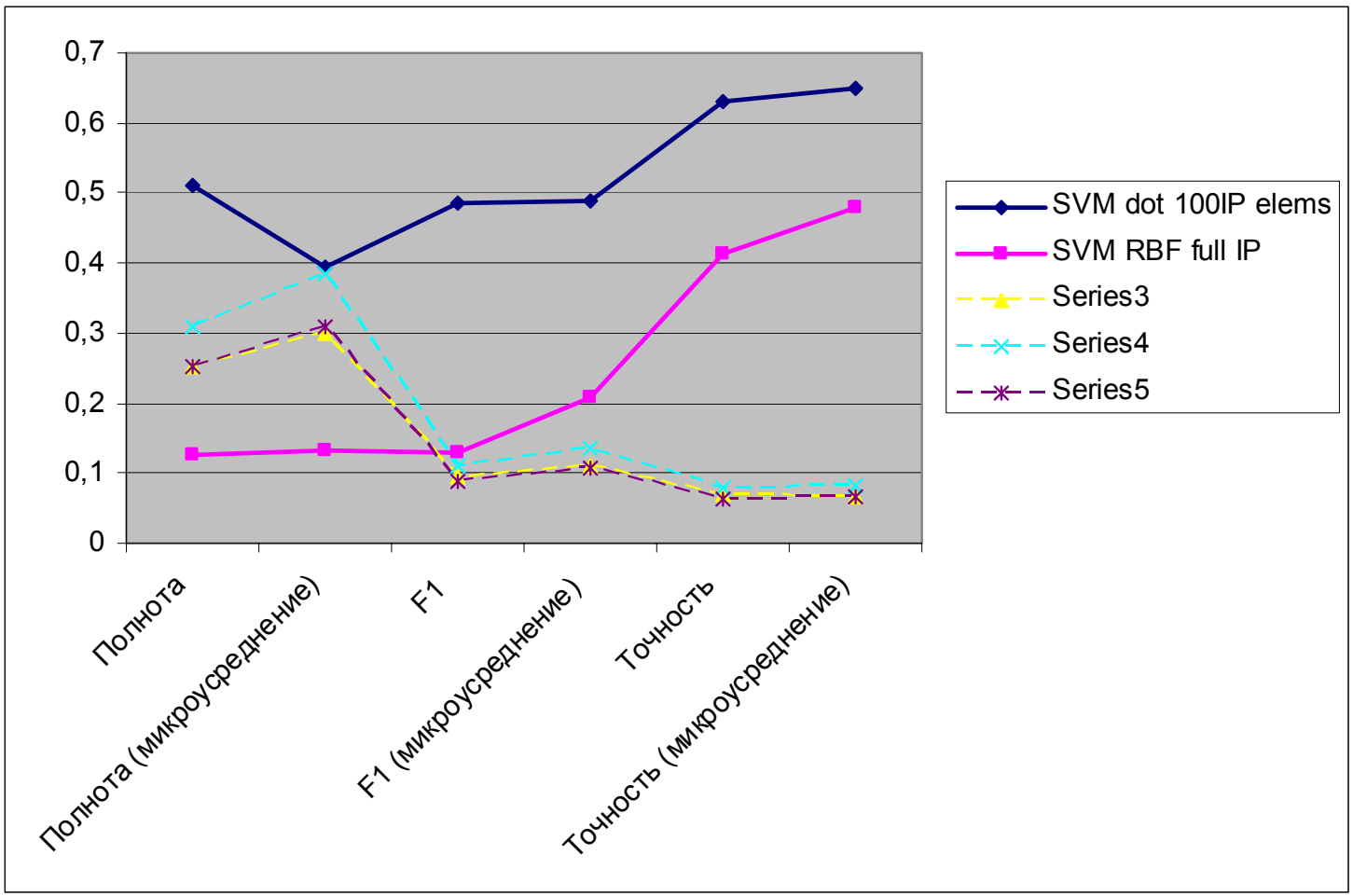
Использование алгоритма кластеризации LSA для задачи классификации документов



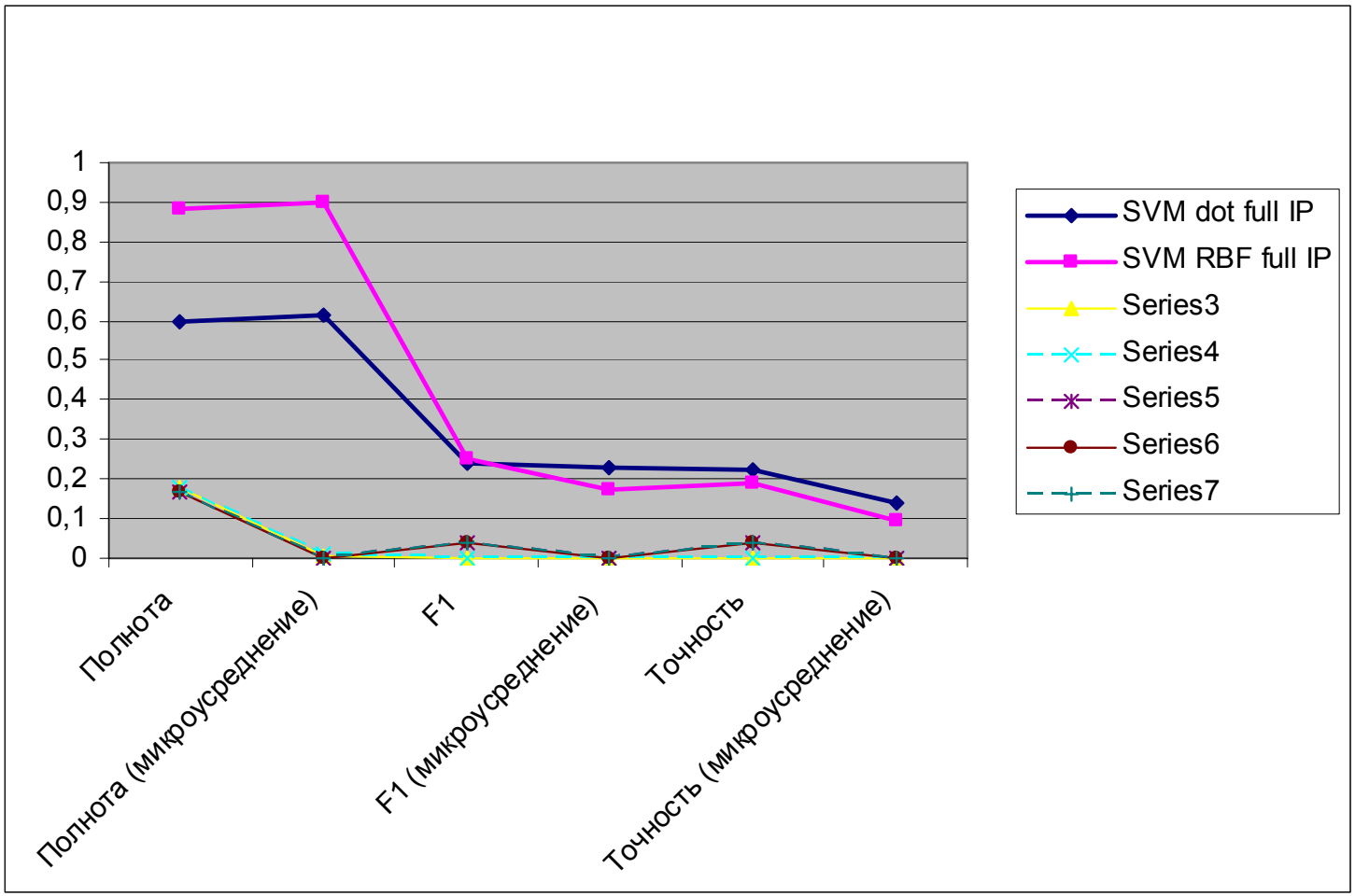
Классификация Веб-сайтов “сильная” оценка



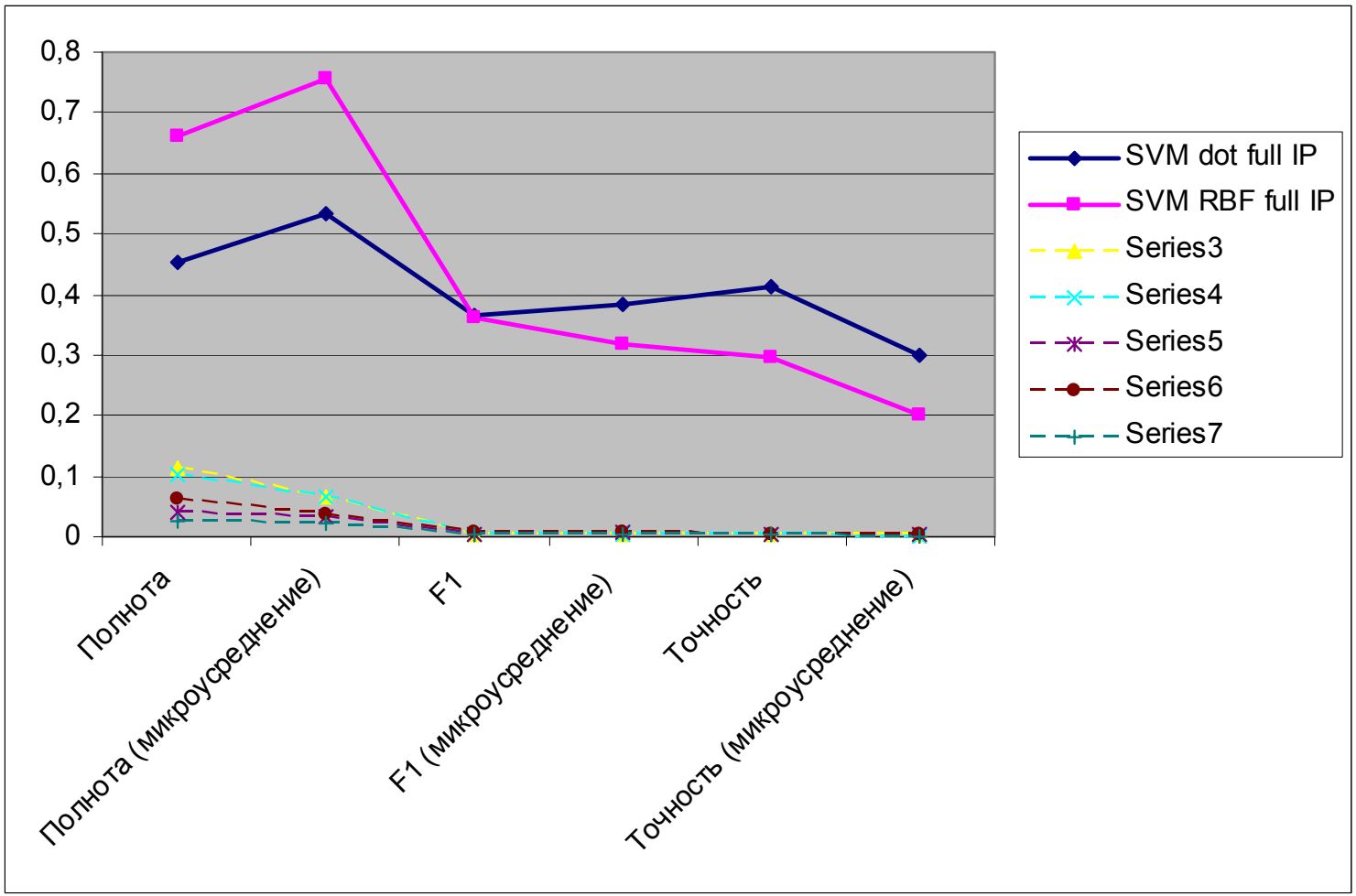
Классификация Веб-сайтов “слабая” оценка



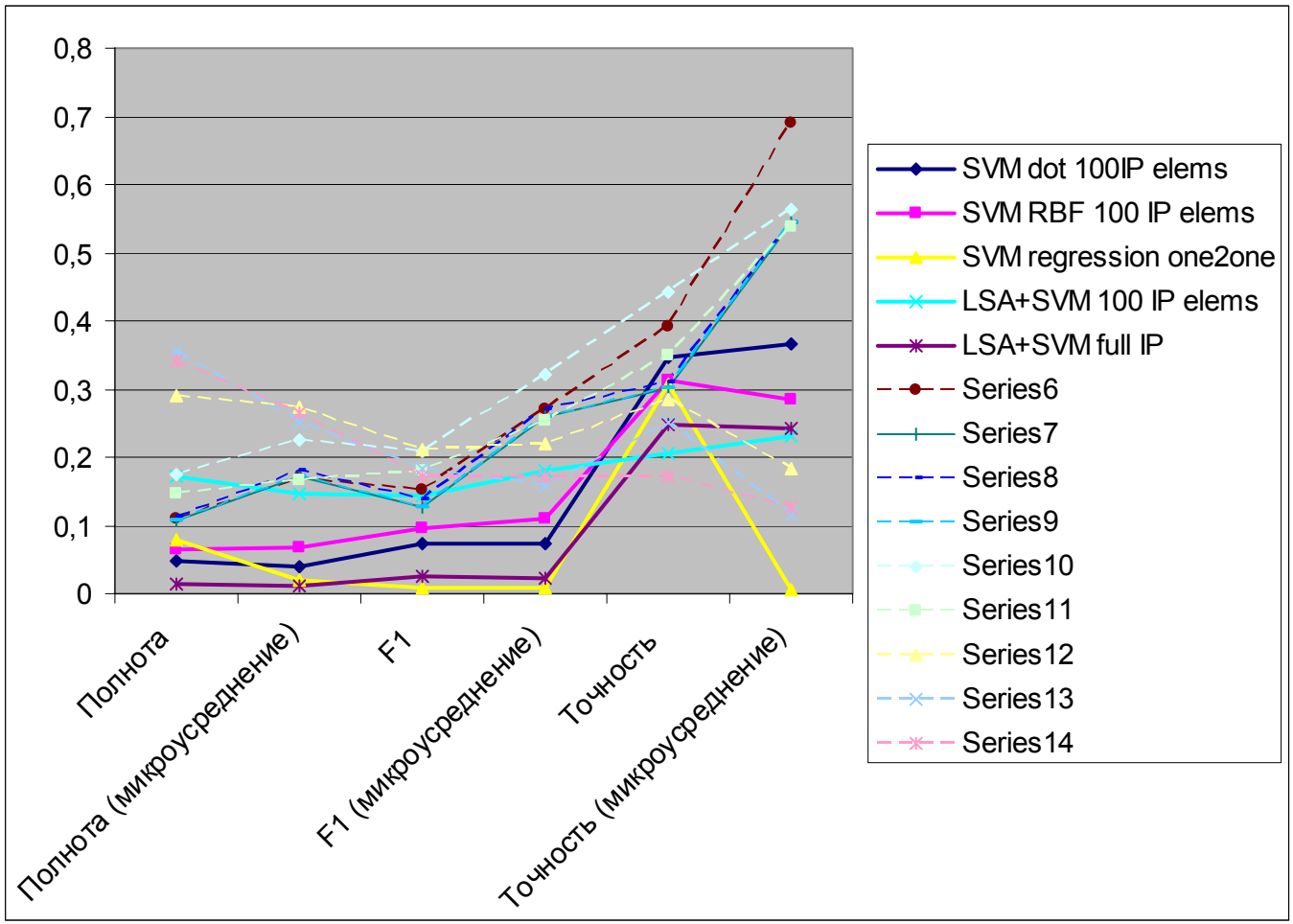
Классификация Веб-страниц “сильная” оценка



Классификация Веб-страниц “слабая” оценка



Классификация нормативно-правовых документов



Выводы

- Подход,сочетающий метод построения ИнфоПортрета с методом опорных векторов, оправдал себя – улучшены показатели классификации по сравнению с результатами прошлого года
- Исследована эффективность сочетаний подходов на различных массивах

Спасибо за внимание

А.В. Антонов,
С.Г. Баглей,
В.С. Мешков
{alexa, baglei, meshkov}
@galaktika.ru