

# Галактика Zoom на РОМИП 2005: оценка модификации метода формирования инфопортрета

---

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



# Поисково-аналитическая система Галактика Zoom

---

- предназначена для аналитической обработки больших объемов неструктурированной текстовой информации
  - сочетает возможности классической поисковой системы, системы сбора текстовых данных и системы аналитической обработки информации
-

# Информационный портрет

---

- основное понятие в системе Галактика Zoom
  - представляет собой список языковых инвариантов (слов и словосочетаний), отличающих данную выборку от прочих
-

# Классификация в системе Галактика Zoom

---

- определение меры близости класса к документу (нормативно-правовому, веб-странице, веб-сайту)
-

# “Старая” модификация метода: проблема “зашумления” инфопортрета

---

ФУНКЦИЯ

ФАЙЛ

БАРХАТНЫЙ ПУТЬ

ЗНАЧЕНИЕ

СТРОКА

ОКНО

ПРИЛОЖЕНИЕ

ХОСТИНГ

УКАЗАТЕЛЬ

ПАРАМЕТР

ВОЗВРАЩАТЬ

ОПЕРАЦИОННАЯ СИСТЕМА

ИДЕНТИФИКАТОР

ФАЙЛОВАЯ СИСТЕМА

ПЕРЕМЕННЫЙ

КЛАСС

---

# Новая модификация метода формирования инфопортрета

---

- Этапы фильтрации инфопортрета
    - формирование "мультидокументов" отдельных сайтов;
    - объединение "мультидокументов" в выборку;
    - формирование инфопортрета выборки;
    - отбор элементов инфопортретов. Из инфопортретов сайтов исключаются те элементы, вес которых в инфопортрете выборки "мультидокументов" незначителен.
-

# Инфопортрет рубрики “Компьютеры / Программирование”

---

## **РОМИП 2004**

ФУНКЦИЯ  
ФАЙЛ  
БАРХАТНЫЙ ПУТЬ  
ЗНАЧЕНИЕ  
СТРОКА  
ОКНО  
ПРИЛОЖЕНИЕ  
ХОСТИНГ  
УКАЗАТЕЛЬ  
ПАРАМЕТР  
ВОЗВРАЩАТЬ  
ОПЕРАЦИОННАЯ СИСТЕМА  
ИДЕНТИФИКАТОР  
ФАЙЛОВАЯ СИСТЕМА  
ПЕРЕМЕННЫЙ  
КЛАСС  
СИМВОЛ  
СТАНДАРТНАЯ ФУНКЦИЯ  
БЕСПЛАТНАЯ УСТАНОВКА  
ОБЪЕКТ

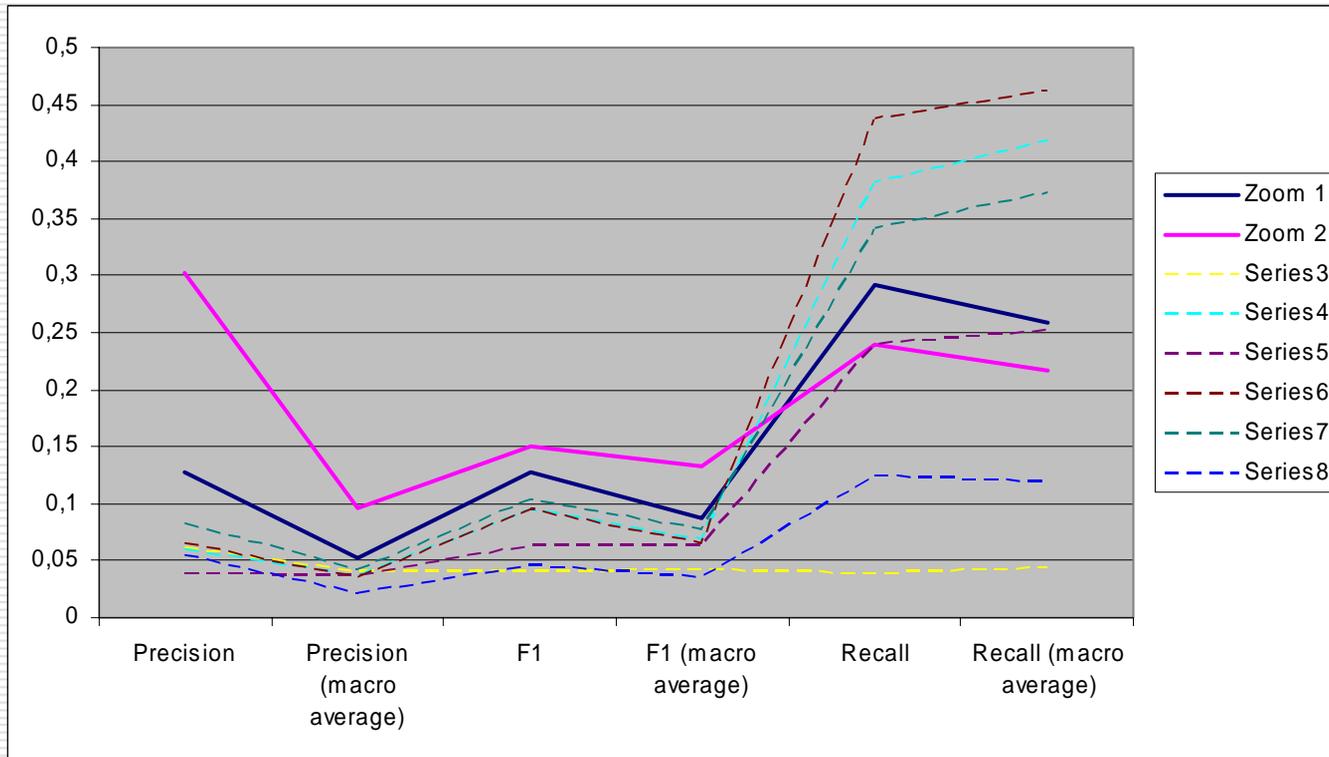
## **РОМИП 2005**

ФАЙЛ  
ФУНКЦИЯ  
ЗНАЧЕНИЕ  
УКАЗАТЕЛЬ  
СТРОКА  
ОКНО  
ПАРАМЕТР  
ВОЗВРАЩАТЬ  
ПЕРЕМЕННЫЙ  
ИДЕНТИФИКАТОР  
ПРИЛОЖЕНИЕ  
ЭТА ФУНКЦИЯ  
ОБЪЕКТ  
ОПЕРАЦИОННАЯ СИСТЕМА  
ПОЛЬЗОВАТЕЛЬ  
СИМВОЛ  
БУФЕР  
ГЛАВНОЕ ОКНО  
КЛАСС  
СООБЩЕНИЕ

---

# Классификация веб-страниц

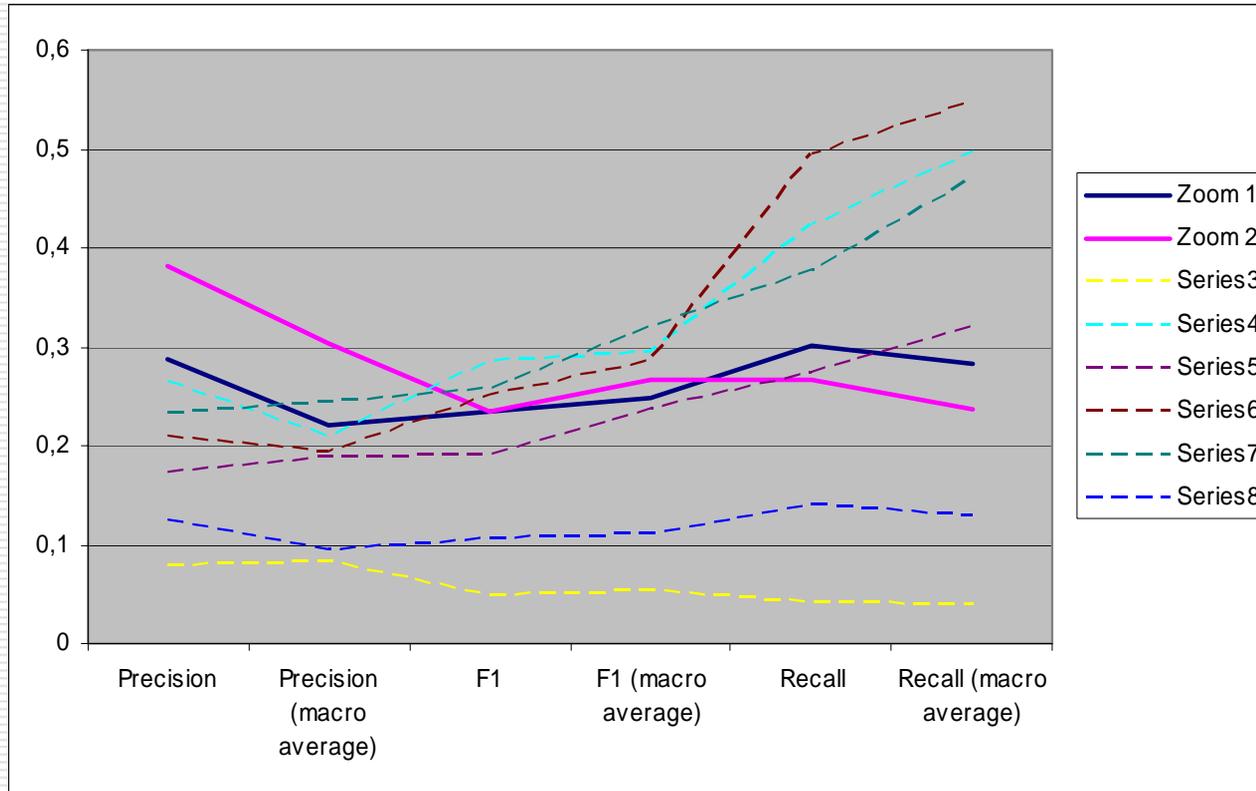
“сильная” оценка, все страницы включены в выборку



# Классификация веб-страниц

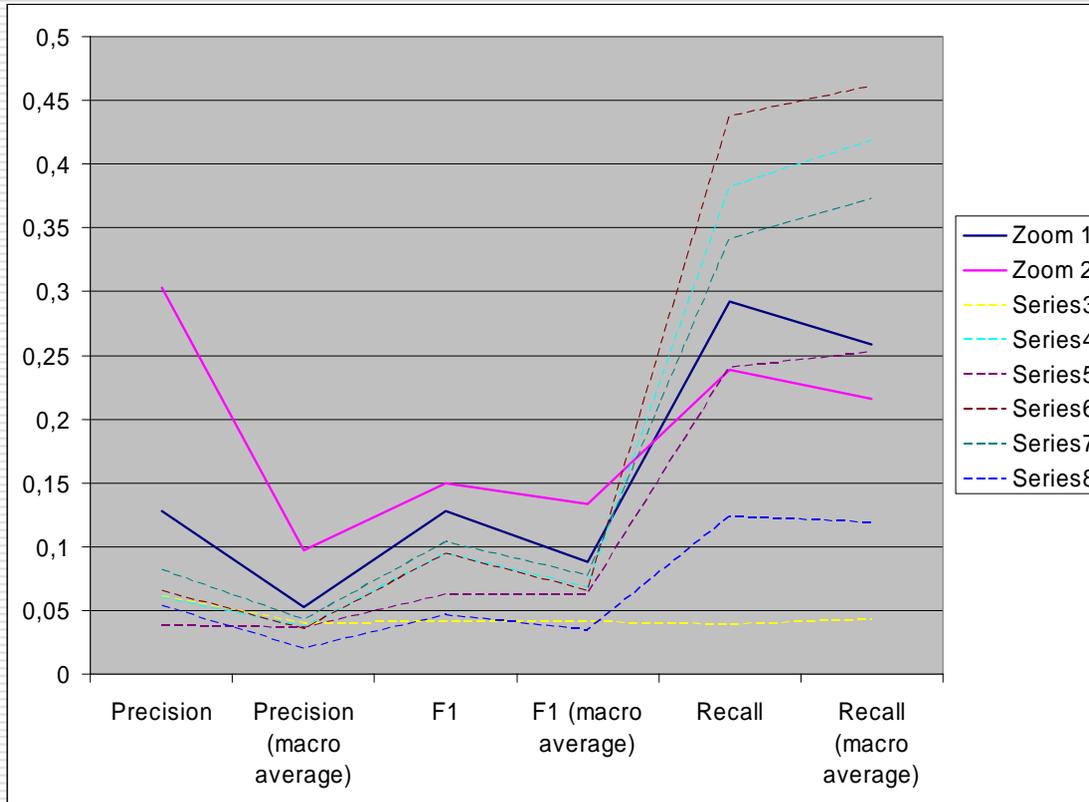
“сильная” оценка, в выборку включены только оцениваемые страницы

---



# Классификация веб-страниц

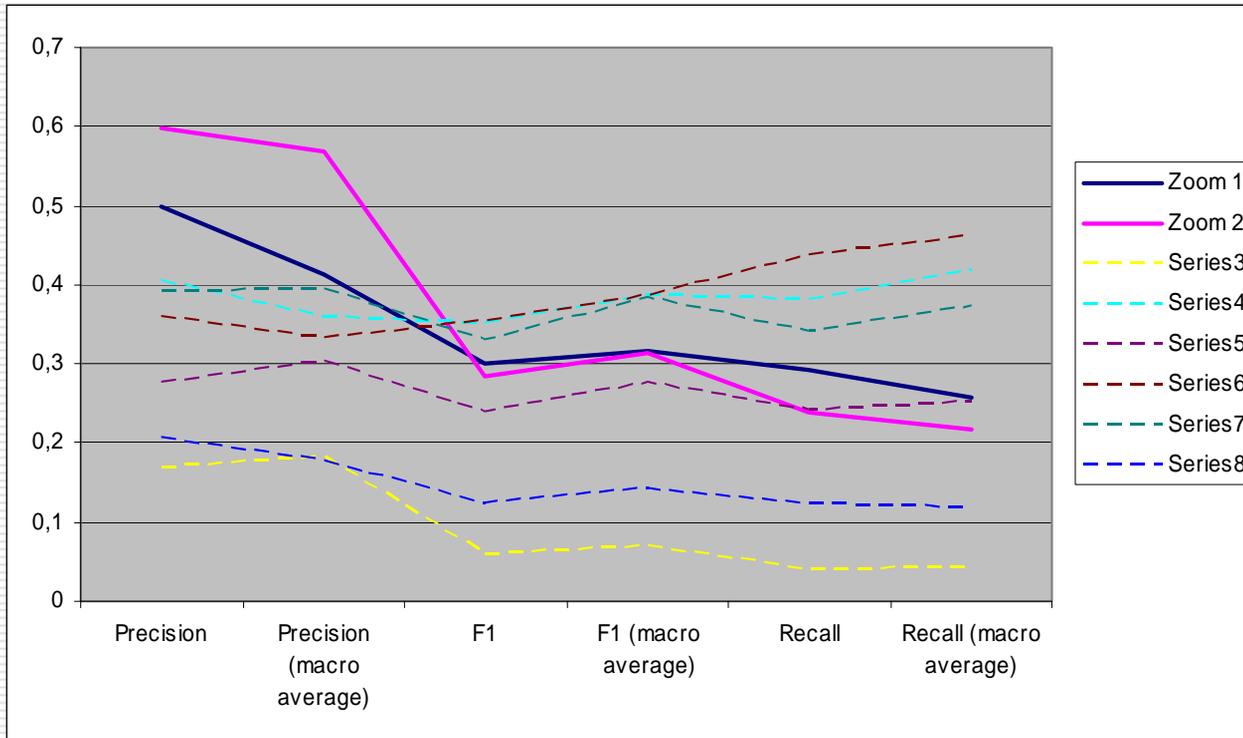
“слабая” оценка, все страницы включены в выборку



# Классификация веб-страниц

“слабая” оценка, в выборку включены только оцениваемые страницы

---



# Инфопортрет рубрики “Общество / Религия и духовность”

---

## □ РОМИП 2004

КН  
СМИРНОВ  
ПОПОВ  
ИВАН  
ПОП  
ИВАНОВ  
БОГ  
ИВАНОВИЧ  
СОЛОВЬЕВ  
ВАСИЛЬЕВИЧ  
НИКОЛАЕВИЧ  
АЛЕКСАНДРОВИЧ  
ЦЕРКОВЬ  
АЛЕКСЕЕВИЧ  
МИХАЙЛОВИЧ  
БОЖИЙ  
ХРИСТОС  
НИКОЛАЙ  
ПЕТРОВИЧ  
ИИСУС

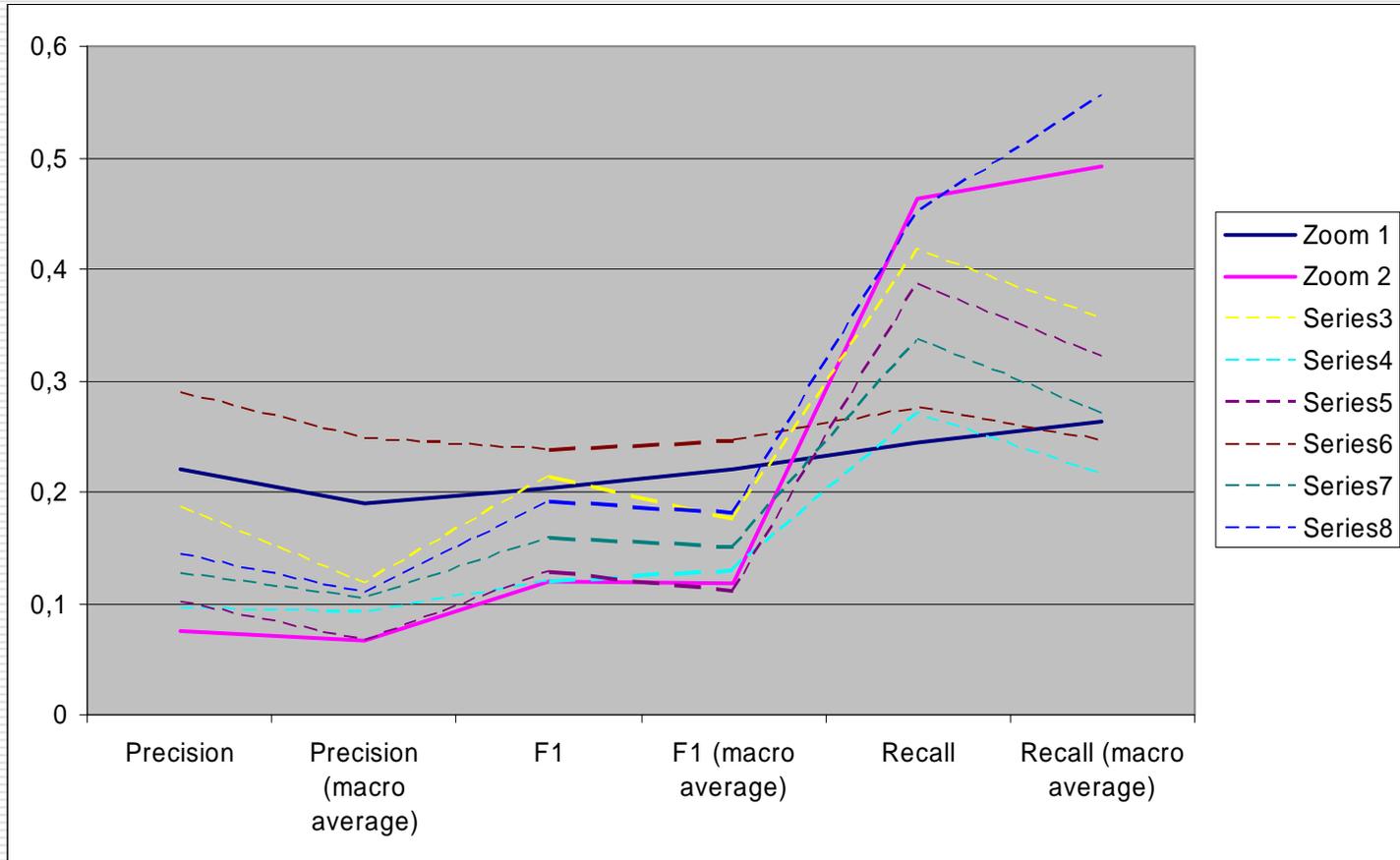
## □ РОМИП 2005

ЦЕРКОВЬ  
БОГ  
ХРИСТОС  
БОЖИЙ  
ГОСПОДЬ  
ИИСУС  
СЛУЖЕНИЕ  
СВЯТОЙ ДУХ  
БИБЛИЯ  
СВЯТОЙ  
ГРЕХ  
МОЛИТВА  
ОТЕЦ  
ПРАВОСЛАВНЫЙ  
СЛУЖИТЕЛЬ  
ПИСАНИЕ  
ПАСТОР  
ПАВЕЛ  
МОЛИТЬСЯ  
ДУХОВНЫЙ

---

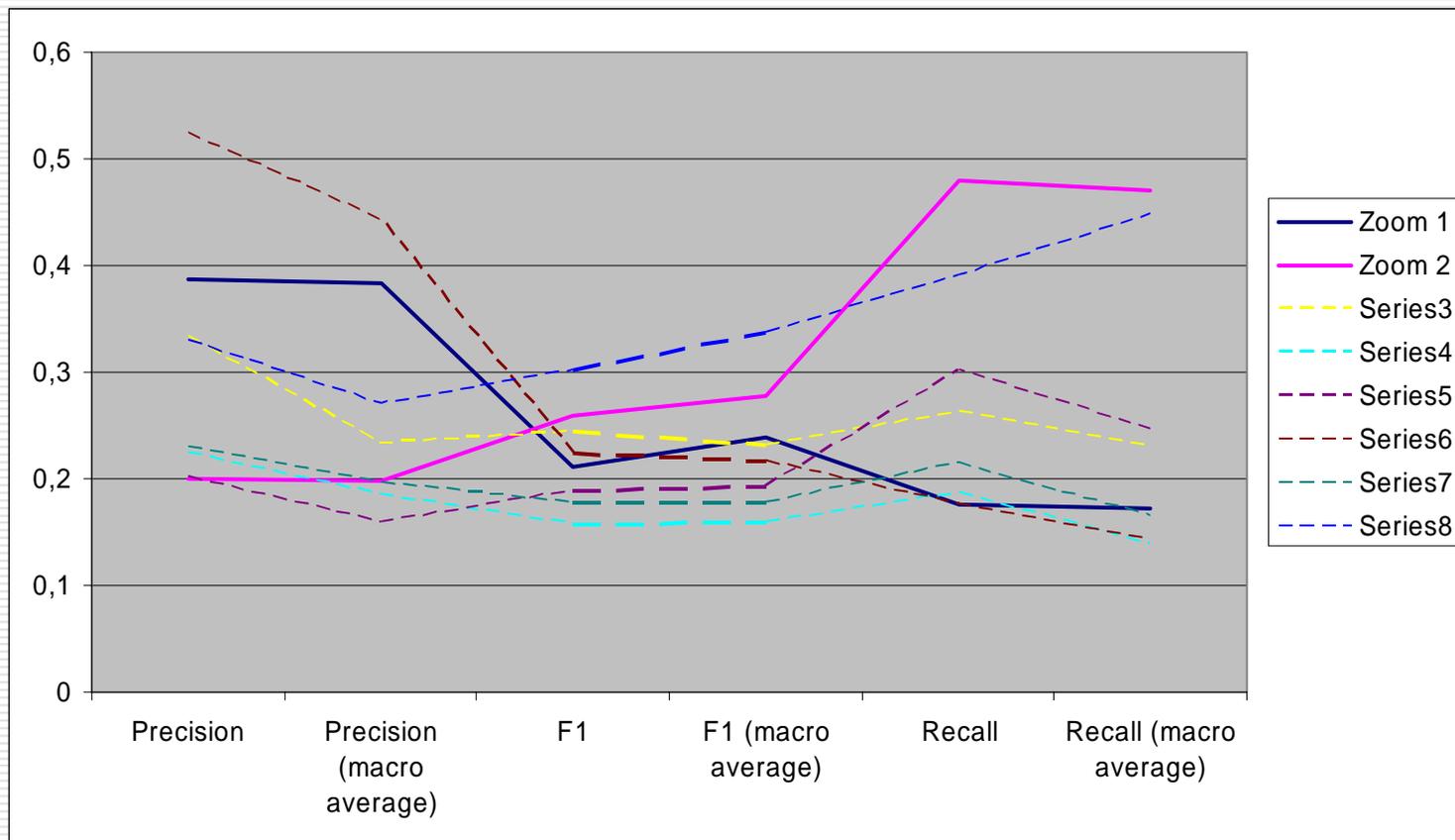
# Классификация веб-сайтов

“сильная” оценка



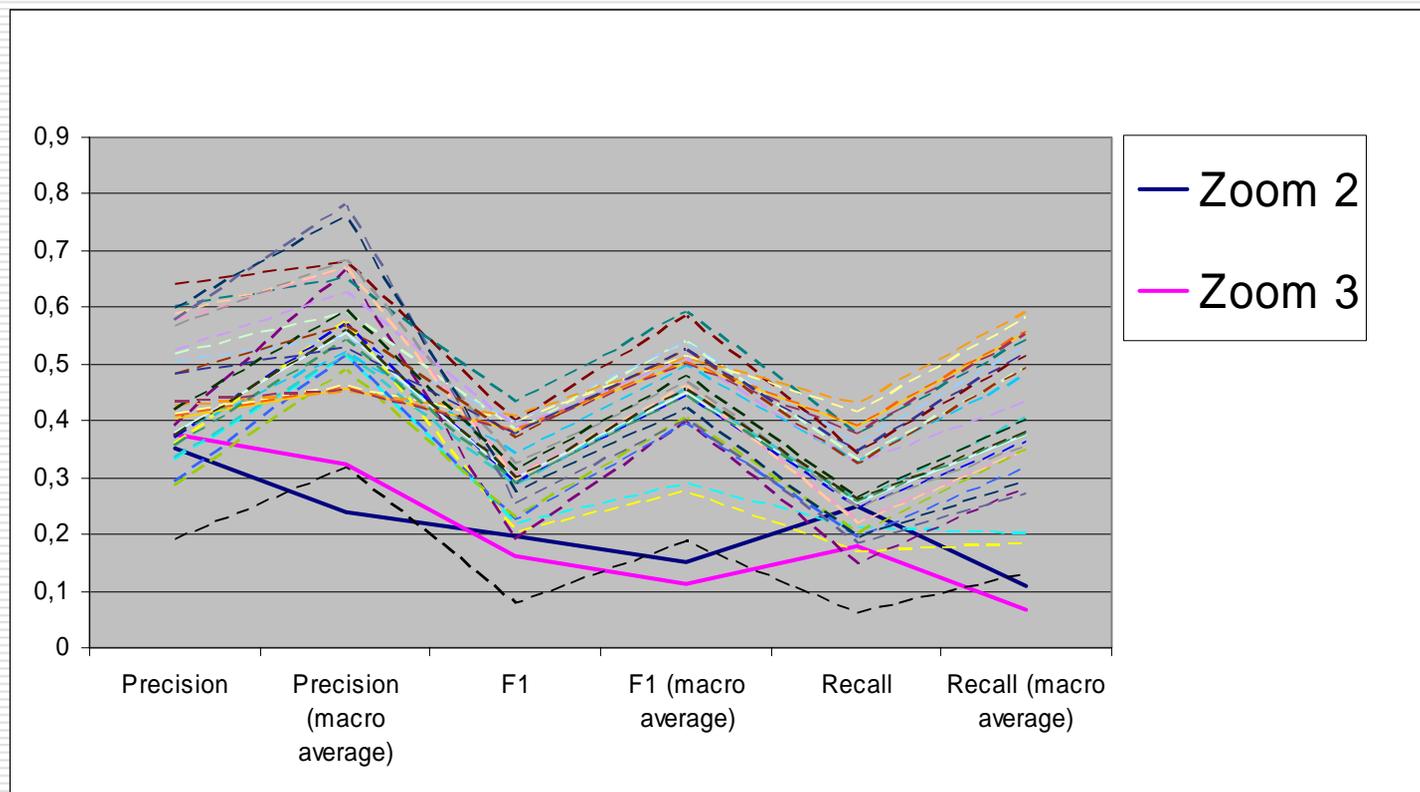
# Классификация веб-сайтов

“слабая” оценка



# Классификация нормативно-правовых документов

---



# Новая модификация метода формирования инфопортрета

---

- Улучшила качество классификации по дорожке "Классификация веб-сайтов"
  - Позволила участвовать в дорожке "Классификация веб-страниц"
-

# Спасибо за внимание

---



<http://zoom.galaktika.ru>