

1. Классы

- Автоновости
- В мире
- В России
- Инопресса
- Культура
- Медицина
- Недвижимость
- Спорт
- Технологии
- Экономика

2. Вектор признаков

Сумка слов, связанных с темой текстов обучающей выборки

3. Способ векторизации

TFIDF

4. Обучающая выборка

Составлена из 50 статей для каждой категории, которая так же является и тестовой выборкой. Для выбора статей был написан запрос к БД следующего вида и представлен в листинге 1.

```
SELECT    id    FROM    XMLCollection_article    where  
category='category' limit 50;
```

Листинг 1 – выборка статей из БД

Где параметр «category» соответствует одному из классов, указанных в пункте 1.