

Метод определения эмоций в текстах на русском языке

КОМПАНИЯ «АЙ-ТЕКО»

2011





Характеристики системы:

1. практическая реализация системы для определения тональности;
2. тексты СМИ на русском языке;
3. обработка грамматически правильных текстов;
4. словари лексической тональности;
5. набор комбинаторных правил объединения отдельных слов и словосочетаний;
6. использование предикационных отношений в пропозиции.

План доклада:

1. Термины и понятия тональности.
2. Описание работы системы.
3. Оценка результата работы модуля определения тональности.



Тональность (сентимент) текста – эмоциональная оценка, выраженная в тексте.

Эмоциональная оценка: позитив vs. негатив (хорошо-плохо).

Пример негативного предложения:

(1) Причиной аварии **Tu-154** стал отказ двигателей.

Пример позитивного предложения:

(2) **Пилоты** смогли посадить машину.

+ **сила тональности** – оценка яркости эмоциональной составляющей слова или высказывания (например, по такой шкале: «сильный негатив, негатив, нейтрально, позитив, сильный позитив»).

Пример сильного негатива:

(3) Состояние здоровья **артиста** сильно ухудшилось.



Лексическая тональность – эмоциональная составляющая, выраженная на уровне лексемы или коммуникативного фрагмента.

Тональность текста в целом определяется лексической тональностью составляющих его единиц и правилами их сочетания.

Эмоциональная оценка в тексте выражается по отношению к **объекту тональности**. Может быть задан заранее или определяться в предложениях как любое имя собственное или даже нарицательное: название компании, название продукта компании, имя или фамилия, географическое название и пр.

Субъект тональности – носитель выраженной в тексте эмоциональной оценки. В случае цитирования, прямой или косвенной речи совпадает с автором высказывания, иначе – с автором текста.

Термины и понятия: три компонента тональности



Тональность определяется тремя компонентами:

- ✓ **Субъект тональности**
- ✓ **Собственно тональная оценка (позитив / нейтрально / негатив, плюс сила тональности)**
- ✓ **Объект тональности**

Например, в (1) субъект тональности – автор текста, объект тональности (определён автоматически) – Ту-154, выражена негативная оценка.

(1) Причиной аварии Ту-154 стал отказ двигателей.

Описание системы: этапы работы



Текст: *Альянс не собирается вмешиваться в ливийский конфликт.*

Морфоанализатор +
тональные словари

Этап 1: разметка текста

альянс<noun,ppos> *не*<invertor> *собираться*
<verb,lnk> *вмешиваться*<verb,negr> *в*<prep>
ливийский<adj,neut> *конфликт*<noun,neg>.

Синтанизатор +
набор правил 1

Этап 2: объединение в цепочки

{*альянс*}<nPPos> {*не собираться вмешиваться*}
<vPosP> {*ливийский конфликт*}<nNeg>.

Выделение объекта
тональности +
набор правил 2

Этап 3: сентимент объекта тональности

{*альянс*}<Pos> *не собираться вмешиваться*
ливийский конфликт.

Результат: {*Альянс не собирается вмешиваться в ливийский конфликт.*}<POS>



Компоненты системы:

- ✓ словари лексической тональности;
- ✓ правила объединения цепочек;
- ✓ правила определения сентимента объекта тональности в предложении;
- ✓ (лингвистический модуль: морфология, синтаксический анализ, определение роли сущности в предложении и её семантического типа).



Тональные словари (составлялись экспертно):

- ✓ прилагательные (позитивные, негативные и усилительные),
- ✓ наречия (позитивные, негативные и усилительные),
- ✓ существительные (позитивные, негативные, потенциально позитивные и потенциально негативные, отглагольные негативные и отглагольные позитивные),
- ✓ глаголы (см. ниже),
- ✓ глагольные коллокации (позитивные и негативные),
- ✓ неглагольные коллокации (позитивные, негативные и усилительные).

Описание системы: глагольные словари



Любое упоминание объекта в предложении характеризуется его окружением и его ролью. Отсюда **8 тональных классов глаголов**:

- ✓ 1 и 2 класс — негативные и позитивные глаголы, определяющие тональность объекта в зависимости от окружения и независимо от роли (негативные *уносить, освободить от*; позитивные *защищать, болеть за*);
- ✓ 3 и 4 класс — негативные и позитивные глаголы, определяющие тональность объекта независимо от окружения, но в зависимости от его роли (*сдаться и проиграть* — субъект негативный, объект позитивный; *обуздать и повергнуть* — субъект позитивный, объект негативный);
- ✓ 5 и 6 класс — негативные и позитивные глаголы, определяющие тональность объекта в зависимости от окружения и роли («возвратные»: негативные *жаловаться, испугаться*; позитивные *окупаться, согреться*);
- ✓ 7 и 8 класс — чисто негативные / позитивные глаголы, определяющие тональность объекта вне зависимости от его роли и окружения (позитивные *расследовать и улучшать*, негативные *грабить и злоупотреблять*);
- + 9 класс — глаголы, приравнивающие тональность объекта и субъекта (т.н. связочные: *являться, олицетворять, относиться*).



Последовательность объединения

- ✓ Сложные предложения разбиваются на простые, некоторые типы придаточных предложений включаются в родительское, причастные обороты присоединяются к определяемому слову, деепричастные — к субъекту родительского предложения.
- ✓ Придаточные предложения определительного типа с разрешенной анафорией соотносятся с определяемым словом.
- ✓ Предложение приводится к одному из типов синтаксической структуры из субъекта, предиката и объекта, где каждый член структуры в общем случае представлен цепочкой словоформ с определенной тональностью.



Последовательность определения сентимента объекта тональности

- ✓ Выделяется объект тональности (может быть задан или выбирается автоматически).
- ✓ Объекту тональности приписывается сентимент на основании типа глагола, а также роли объекта тональности и его позиции в предложении.

Пример:

Милиционер	подорвался на	бомбе.	
Subj, Inq, nPPos	vNegP	nNeg	= nNeg

Милиционер	изъял	бомбу.	
Subj, Inq, nPPos	vNeg	nNeg	= nPos

- ✓ Сентимент и сила сентимента объекта тональности подсчитываются по предложению и по всему тексту.



Тестирование системы:

- ✓ тестовый модуль в сети (<http://x-file.su/tm/>), отслеживание логов;
- ✓ периодические субъективные оценки небольших текстовых подборок экспертом:
 - один раз в неделю,
 - первые 5-7 новостных текстов с rbc.ru за понедельник или вторник,
 - в среднем по 70 предложений в неделю,
 - 1121 предложение с января по начало мая 2011 г.

Оценка результата работы: периодическое тестирование



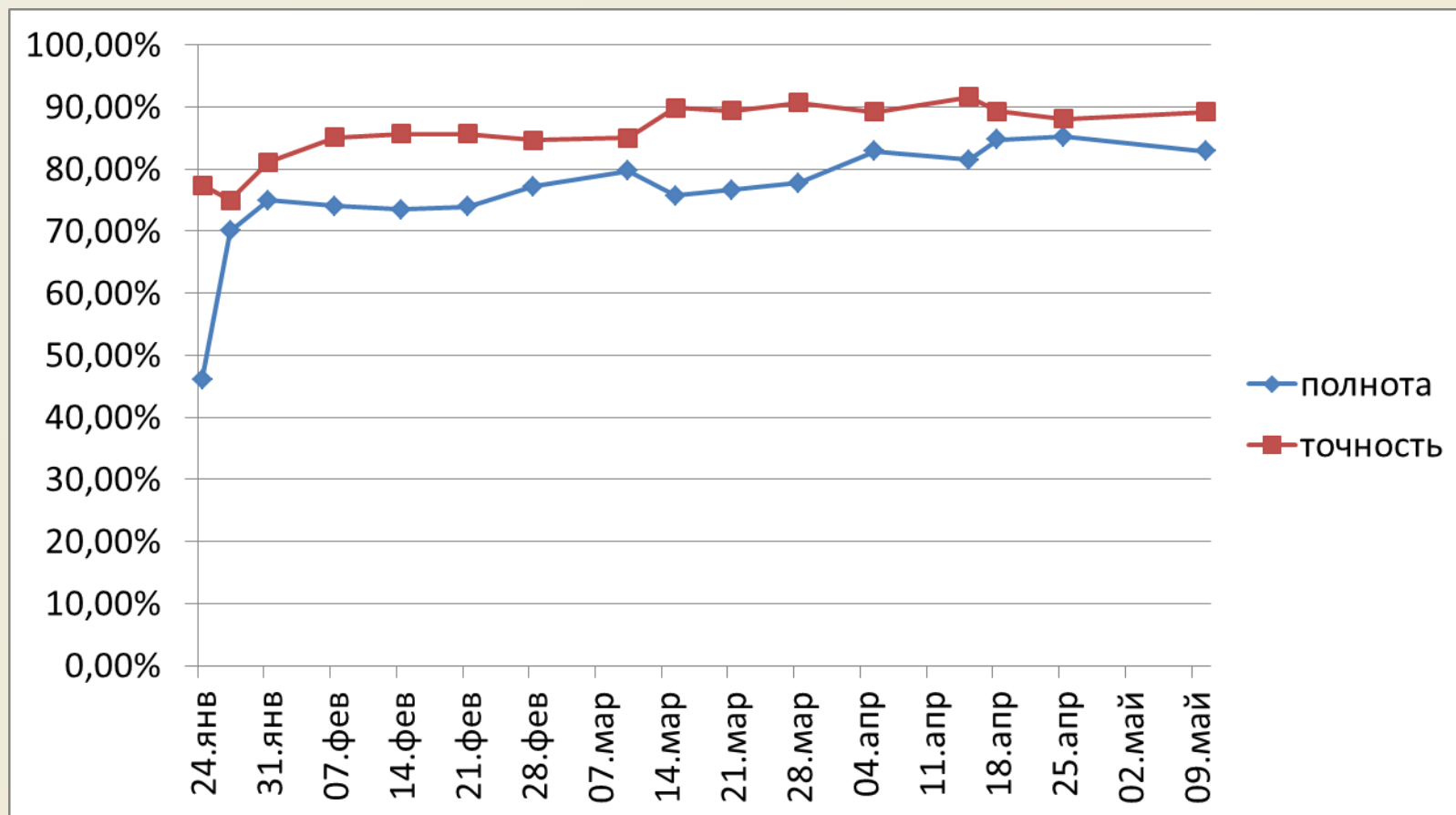
№	предложение	ОК (A)	пропуск (B)	знак (C)	лишнее (D)
42	В результате взрыва на АЭС "Фукусима-1 " поврежден реактор.			1	
43	На четвертом реакторе АЭС "Фукусима-1 " в 11: 53 по местному времени(05: 53 мск) произошел взрыв водорода, передают японские СМИ.	1			
44	В 11: 14 по местному времени(05: 00 мск) в зоне четвертого реактора начался пожар, сообщили в компании-операторе станции Tokyo Electric Power(ТЕРСО).		1		

- ✓ тональных предикаций в тексте: $A + B + C$
- ✓ предикаций, определённых системой как тональные: $A + C + D$
- ✓ полнота: $A / (A + B + C)$, точность: $A / (A + C + D)$

Оценка результата работы: изменение качества



Изменение качества тональной разметки с января по март 2011 г.



Текущее качество тональной разметки – около 83% полноты, 90% точности



При определении тональности возникают три класса ошибок:

- ✓ ошибки работы модуля морфологической и синтаксической разметки текста (около 5-7%);
- ✓ ошибки правил комбинаторики (не более 3%);
- ✓ ошибки тональных словарей, вызванные их неполнотой и «тональной» омонимией (не более 5%).

Кроме того:

- ✓ ограниченность используемого эмотивного пространства;
- ✓ недостаточная исследованность вопросов размерности и формализации спектра человеческих эмоций.

**Спасибо за
внимание**

<http://x-file.su/tm/>