



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Исследование возможности оценки свободных ответов на естественном языке по ключевым словам

А. П. Савинов





О докладчике

- Анатолий Павлович Савинов
- Доцент Национального исследовательского Томского политехнического университета
- Ведущий менеджер института Электронного образования ТПУ.
- Разработчик обучающих систем с диалогом на естественном языке



Тесты №1 и №2 демонстрирующие проблемы, возникающие при работе ИОС

- 1. Перечислите обобщающие показатели использования основных фондов.***

Эталонный ответ: Фондоотдача, фондоемкость, фондовооруженность.

- 2. Дайте определение термину «амортизация».***

Эталонный ответ: Амортизация – это планомерный процесс переноса стоимости средств труда по мере их износа на производимый с их помощью продукт.



Продолжение: *Тест № 3*

3. *Дайте определение термину «фонды обращения»*

Эталонный ответ: Фонды обращения- это оборотные средства, обслуживающие процесс реализации готовой продукции; служат для обеспечения непрерывности процесса производства и реализации продукции предприятия (примеры: готовая продукция на складе, товары, отгруженные заказчикам, но еще не оплаченные ими, дебиторская задолженность, средства в расчетах, денежные средства в кассе предприятия и на счетах в банках).



Продолжение: *Тест №4*

4. По каким объектам основных фондов амортизация не начисляется?

Эталонный ответ: Амортизация не начисляется по следующим объектам основных средств: объектам, полученным по договору дарения и безвозмездно в процессе приватизации; жилищному фонду (кроме объектов, используемых для извлечения дохода) объектам, потребительские свойства которых с течением времени не изменяются.



Алгоритм оценки свободных ответов на ЕЯ

Оценка ответов осуществлялась по следующему алгоритму:

1. Анализ текста эталона и ответа студента с помощью **АВВУ Compreno**. Результат - синтаксическое дерево.
2. Выделяются ключевые слова и словосочетания.
3. Они фильтруются с помощью **Википедии и Рутеза**.
4. Рассчитывается значение семантической близости между терминами эталона и ответа студента.

5. Конечная оценка ответа вычисляется по формуле:

$$M = \frac{\sum_{i=1}^n r(t_i, s_i)}{|T|},$$

где T – это множество терминов, выделенных из эталона.

s_i – термин, выделенный из ответа студента, стоящий в паре с термином t_i в эталоне;

$r(t_i, s_i)$ - значение семантическая близости.



Оценка семантической близости терминов

Для этих целей на основе **Википедии** и тезуруса Рутез были сделаны две графовых базы данных.

В узлах графа, построенного по **Википедии**, находились заголовки статей. Рёбра — ссылки между статьями с указанием их типа: обычная ссылка.

Семантическая близость

по формуле **Дайса**:

$$r = \frac{2|N(a) \cap N(b)|}{|N(a)| + |N(b)|}$$

$N(a)$ - множество статей, на которые ссылается статья термин « a ».

$N(b)$ - множество статей, на которые ссылается статья термин « b ».



Оценка семантической близости терминов

В узлах графа, построенного **по Рутезу**, находились все текстовые входы тезауруса, а в качестве рёбер использовались связи «выше – ниже».

Семантическая близость по Рутезу определялась в соответствии со следующим алгоритмом:

- Рассчитывается глубина (расстояние от корня) ближайшего предка обоих терминов.
- Если общий предок найден,

близость вычисляется по формуле:

$$r = \frac{2p_c}{p_1 + p_2},$$

где p_c – путь от корня до общего предка,

p_1 – путь от корня до первого термина.

p_2 – путь от корня до второго термина.



Качество автоматической оценки ответов

Оценивалось на 112 студентах по следующим критериям

1. **Процент совпадения ответов,**
оценённых и машиной и
преподавателем, как **правильные:**

$$\epsilon_C = \frac{|A_{C,T} \cap A_{C,M}|}{|A_{C,T}|} \cdot 100\%,$$

2. **Процент совпадения ответов,**
оцененных и машиной и
преподавателем, как **неправильные:**

$$\epsilon_W = \frac{|A_{W,T} \cap A_{W,M}|}{|A_{W,T}|} \cdot 100\%,$$

3. **Общий процент совпа-**
дения оценок машины и преподавателя:

$$\epsilon = \frac{|A_{C,T} \cap A_{C,M}| + |A_{W,T} \cap A_{W,M}|}{|A|} \cdot 100\%,$$

A_{CT} – множество ответов, оцененных преподавателем, как правильные; A – все множество ответов; A_{CM} – множество ответов, оцененных машиной, как правильные;



Таблица с результатами обработки ответов

	ID студента	Оценка преп.	Оценка сист.	Близость
Вопрос: К обобщающим показателям использования основных фондов относятся				
Эталон: фондоотдача, фондо емкость, фондовооруженность.				
Процент совпадения: 98,1				
Процент совпадения верных ответов: 100				
Процент совпадения неверных ответов: 96,5				
фондо емкость, фондоотдача, фондовооруженность	118	1	1	Расчёт
фондо емкость	130	0	0,67	Расчёт
коэффициент износа, коэффициент оборота, фондо ёмкость, фондовооруженность.	148	1	0,67	Расчёт



Пример выделения терминов из текста

Электрическая_емкость
Вместимость_(параметр)
Емкий_(краткий_и_содержательный)
Сосуд_(вместилище)
Емкость
Электрическая емкость
Емкость (электрическая)

OK



Пример расчета семантической близости

[Фондо] ↔ [Фондо] : 1

[Емкость] ↔ [Емкость] : 1

[Фондоотдача] ↔ [Козффициент износа] : 0

Оценка: 0,6666666666666667

OK



Анализ результатов: *Тест №1*

$e = 98,1 \%$; $ec = 100\%$; $ew = 96,6 \%$;

1. В оценках машины **присутствуют искажения**, обусловленные неправильным определением терминов в эталоне и ответе студента с помощью Википедии и Рутеза.
2. Из трёх терминов ими был **правильно выделен** только термин **фондоотдача**.
3. Термин **фондоёмкость** был **заменён** данными системами на два термина **фондо** и **ёмкость**, а термин **фондовооружённость** в базах Википедии и Рутеза вообще **отсутствует**. Результат - **завышение оценки**.
4. **Вывод:** использование баз Википедии и Рутеза недостаточно для выделения ключевых слов и вычисления их семантической близости. ***Необходима онтология «Экономики предприятия»***



Анализ результатов: *Тест №2*

$$e = 84,5\%; \quad e_c = 81\%; \quad e_w = 89,7\%;$$

1. В 8 ответах машинная оценка была завышена (система поставила «правильно» ответу, оценённому преподавателем, как неправильный), а в 1 занижена.
2. Один из 8 вариантов ответа: «Амортизация — **планомерный процесс переноса стоимости продукции объектам основных фондов.**»
3. **Смысл**, вложенный студентом в данное предложение, полностью **противоречит смыслу**, заложенному в эталоне.
4. Система не реагирует на порядок следования ключевых слов в ответе. На оценку влияет только их количество.
5. **Вывод: необходимо учитывать не только состав ключевых слов в тексте, но и взаимную связь между ними.**



Анализ результатов: *Тест №3*

$$e = 54,1\%; \quad e_c = 0\%; \quad e_w = 100\%;$$

1. Преподаватель, чтобы обеспечить гибкость эталонному ответу, **вложил** в него **избыточную информацию**.
2. Ни один ответ студента компьютером не оценен, как правильный. Обусловлено тремя причинами:
 - **Дублированием информации**. В эталонном ответе даются **определение и примеры**, что привело к **двойному увеличению** в нём **объёма** ключевых слов, на который производится нормирование интегрального коэффициента семантической близости ключевых слов, содержащихся в тексте ответа и в эталоне.
 - Стремлением студентов **кратко излагать** свои мысли в ответах на поставленный вопрос.
 - Преподавателей **устраивает краткий ответ** студентов.



Анализ результатов: *Тест №4*

$$e = 75,4\%; \quad e_c = 12,5\%; \quad e_w = 100\%;$$

1. **Результаты** статистического анализа *Теста №4* близки по величине *Тесту №3*.
2. Наблюдаемое занижение оценки компьютером в *Тестах №3 и №4*, обусловлено разными причинами:
 - В *Тесте №3* было дублирование информации, а в *Тесте №4* в эталон была введена неполная информация о том в каких случаях амортизация не начисляется.
 - Полный ответ на этот вопрос на **75% шире**, чем дан в эталоне преподавателем.
 - Студенты **излагают дополнительные** по отношению к эталону **сведения**, полученные из других учебников, а преподаватели, располагая большим багажом знаний, **положительно оценивают** эти ответы.



Алгоритм формирования и работы с эталоном

1. Эталонный ответ должен содержать полную релевантную вопросу информацию без дублирования.
2. Преподаватель разбивает эталон на совокупность компонентов, каждый из которых выражает одну законченную мысль. Каждый компонент оценивается им с точки зрения важности.
3. Система автоматически определяет значимость ключевых слов в компонентах эталона и в ответе студента.
4. В случае неполного ответа студента система генерирует дополнительные вопросы по компонентам эталона, которые не были освещены в ответе.



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Спасибо за внимание!

Анатолий Павлович Савинов
savinov@tpu.ru

