

Research on automatic natural language answer assessment method based on bag-of-concepts approach.

A.P. Savinov





About

- Anatoly Pavlovich Savinov
- Tomsk Polytechnic University docent
- TPU e-learning institute top manager.
- Natural language based e-learning systems developer.



Tests №1 and №2 demonstrating intellectual e-learning systems problems (in russian)

- 1. Перечислите обобщающие показатели использования основных фондов. (List the indicator of fixed assets use)*

Correct answer: Фондоотдача, фондоемкость, фондовооруженность.
(capital productivity, capital intensity, capital-labor ratio)

- 2. Дайте определение термину «амортизация». (Give the depreciation definition)*

Correct answer: Амортизация – это планомерный процесс переноса стоимости средств труда по мере их износа на производимый с их помощью продукт. (Depreciation is a process of transferring assets cost to product cost with assets wear and tear.)



Test № 3

3. Дайте определение термину «фонды обращения» (Give the circulating assets definition).

Correct answer: 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).

Circulating assets are current assets serving the process of the finished product realization; serve to ensure the continuity of production and sales of enterprise process (examples: the finished products in the warehouse, the goods shipped to customers, but not yet paid for by them, receivables, funds in the calculations, cash in hand and enterprises on bank accounts).



Test №4

4. *По каким объектам основных фондов амортизация не начисляется?* (Which fixed assets object are not included in the depreciation?)

Correct answer: 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.

Following objects are not included in depreciation: objects obtained under the contract of donation and donated in the privatization process; housing stock (except for objects used to generate income); objects whose consumer properties do not change over time.



Natural language answer assessment algorithm

Following algorithm was used for answer assessment:

1. Answer text and correct text analysis with **ABBYY Compreno** analyzer. Result produced is a syntactic tree.
2. Keywords and collocations search.
3. Keywords filtering using Wikipedia and Ruthes thesaurus (only words containing in these resources are used).
4. Calculation of semantic relatedness between student terms and correct answer terms.

5. Final grade is calculated as follows:
$$M = \frac{\sum_{i=1}^n r(t_i, s_i)}{|T|},$$

where T is a set of correct answer terms.

s_i is a student term paired with correct answer term t_i ;

$r(t_i, s_i)$ is these terms semantic similarity value.



Terms semantic similarity measure

Two graph-based databases were build using Wikipedia and Ruthes thesaurus.

Wikipedia-based graph contains article titles as nodes and links between these articles as edges.

Semantic similarity

Dice metric:

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

$N(a)$ – set of articles which term « a » article has links to.

$N(b)$ – set of articles which term « b » article has links to.



Terms semantic similarity measure

Ruthes-based graph contains thesaurus concepts as nodes and “Subclass of” relations as edges.

Ruthes-based semantic similarity is calculated as follows:

- Depth (distance from tree root) of two terms closest common ancestor is calculated.
- If common ancestor is found, similarity is calculated using formula:

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

where p_c – path from tree root to common ancestor,

p_1 – path from tree root to the first term.

p_2 – path from tree root to the second term.



Automatic answer assessment quality

Estimated based on 112 students answers using following criteria:

1. Right answers grade correlation with experts:

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

2. Wrong answers grade correlation:

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

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

3. All answers grade correlation:

A_{CT} – set of answers marked by expert as right; A – set of all answers; A_{CM} – set of answers marked by system as right;



Table of answers processing results

	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

Concept extraction example

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

OK



Semantic similarity calculation example

[Фондо] ↔ [Фондо] : 1

[Емкость] ↔ [Емкость] : 1

[Фондоотдача] ↔ [Козффициент износа] : 0

Оценка: 0,6666666666666667

OK



Results analysis: Test №1

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

1. System grades contains errors caused by incorrect or incomplete concepts extraction from Wikipedia and Ruthes databases.
2. One of three terms was extracted incorrectly.
3. Complex term **фондоёмкость** was split in two terms **фондо** and **ёмкость** by syntactic analyzer; term **фондовооружённость** is missing in both Ruthes and Wikipedia. As a result system grade was too high.
4. **Conclusion:** additional subject area ontology is required for proper terms extraction along with general purpose resources.



Results analysis: Test №2

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

1. In 8 cases system grade was too high, in 1 case too low.
2. One of the answer variants **«Амортизация — планомерный процесс переноса стоимости продукции объектам основных фондов.»** contradicts correct answer.
3. System counts this answer as correct because the same words are used as in correct answer.
4. Conclusion: not only set of terms is important but also relations between them.



Results analysis: Test №3

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

1. Teacher has included some optional parts in correct answer.
2. None of student answers was graded as correct by system due to following reasons:
 - **Information duplication.** Correct answer contains both subject definition and examples, which leads to its length increase. Length (number of terms) is one of values used for grade calculation.
 - Students tend to answer as short as possible.
 - Teachers are often satisfied with student's short answers.



Results analysis: Test №4

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

1. Statistically Test 4 results are closed to Test 3 results.
2. Grade drop in tests 3 and 4 has different causes:
 - Test 3 correct answer has information duplication, while test 4 correct answer was incomplete.
 - Test 4 complete answer is 75% longer than answer specified by teacher.
 - Students show additional knowledge from sources other than correct answer or textbook. Teachers can assess this knowledge for higher grade while system cannot.



Reference answer composition algorithm

1. Reference answer should be complete and only contain relevant information without duplication.
2. Reference answer should be split into a set of components each reflecting a single complete thought.
3. System must automatically estimate weight of terms in student answer and reference answer.
4. If student answer is incomplete, system should be able to ask additional questions.



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Thanks for your attention!

Anatoly Pavlovich Savinov
savinov@tpu.ru

