CAT and the Kittens

Корпус Хайленда (“Is Academic Writing Becoming More Informal?”)

Table 1
Corpus size and composition.

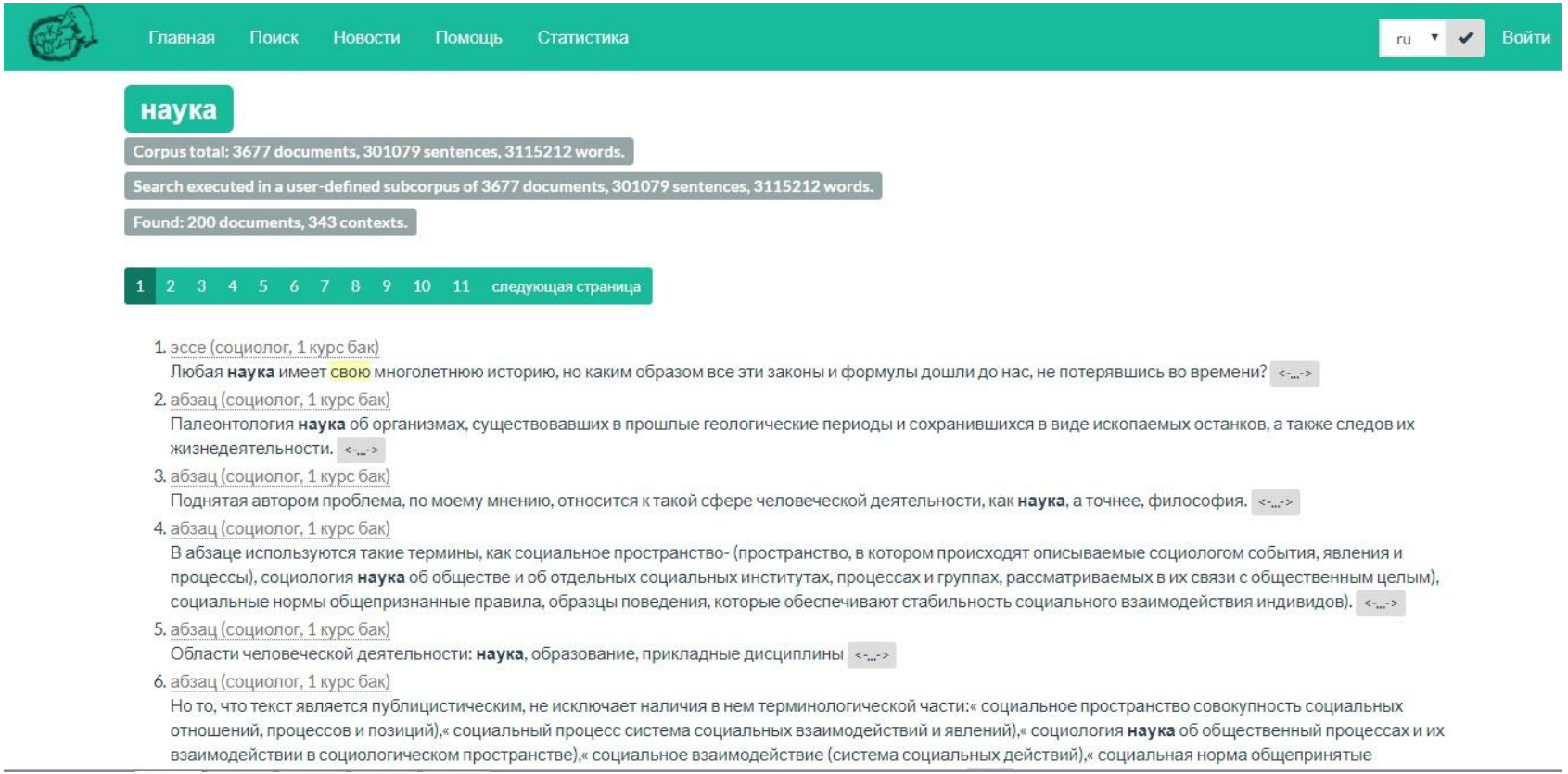
Discipline	1965	1985	2015	Overall
Applied linguistics	110,832	144,859	237,452	493,143
Biology	244,706	263,465	237,998	746,169
Engineering	92,062	97,545,	235,681	425,288
Sociology	149,788	196,232	262,203	608,223
Totals	597,388	604,556	973,334	2,272,823

Is that so?

Changes in use of informality features by discipline (per 10,000 words).

Feature	Applied Ling			Sociology			Electrical			Biology		
	1965	1985	2015	1965	1985	2015	1965	1985	2015	1965	1985	2015
First person	60.9	53.6	52.5	46.5	42.8	68.6	46.9	53.9	68.8	10.9	23.6	34.1
Unattended reference	108.9	70.8	71.8	104.9	75.0	66.2	61.9	60.6	58.9	80.7	62.5	44.1
Initial conjunctions	30.6	32.0	37.9	38.3	37.7	49.0	30.1	31.2	38.6	19.3	29.6	40.1
Second person	4.9	6.4	10.0	9.3	7.2	4.4	0.0	0.2	0.0	0.0	0.2	0.3
Listing expressions	3.5	2.8	1.9	1.6	1.4	1.0	2.3	1.3	1.0	0.4	0.4	0.1
Contractions	1.8	4.8	13.5	2.3	3.3	5.1	0.0	0.1	0.1	0.0	0.1	1.2
Preposition ending	1.3	0.3	1.0	1.2	0.5	0.9	0.2	0.3	0.5	0.0	0.0	0.3
Exclamation	1.3	1.2	0.8	0.4	0.1	1.1	0.0	0.1	0.0	0.3	0.1	0.0
Split infinitives	0.6	0.7	2.1	0.5	0.8	2.5	0.8	2.3	2.4	0.1	1.3	2.3
Direct questions	0.0	0.0	0.2	0.1	0.2	0.1	0.1	0.0	0.1	0.5	0.1	0.1

КРУТ (CoRST)


Главная
Поиск
Новости
Помощь
Статистика
ru
Войти
наука
Corpus total: 3677 documents, 301079 sentences, 3115212 words.
Search executed in a user-defined subcorpus of 3677 documents, 301079 sentences, 3115212 words.
Found: 200 documents, 343 contexts.
1 2 3 4 5 6 7 8 9 10 11 следующая страница
1. эссе (социолог, 1 курс бак)
Любая наука имеет свою многолетнюю историю, но каким образом все эти законы и формулы дошли до нас, не потерявшись во времени? <-...->
2. абзац (социолог, 1 курс бак)
Палеонтология наука об организмах, существовавших в прошлые геологические периоды и сохранившихся в виде ископаемых останков, а также следов их жизнедеятельности. <-...->
3. абзац (социолог, 1 курс бак)
Поднятая автором проблема, по моему мнению, относится к такой сфере человеческой деятельности, как наука, а точнее, философия. <-...->
4. абзац (социолог, 1 курс бак)
В абзаце используются такие термины, как социальное пространство- (пространство, в котором происходят описываемые социологом события, явления и процессы), социология наука об обществе и об отдельных социальных институтах, процессах и группах, рассматриваемых в их связи с общественным целым), социальные нормы общепризнанные правила, образцы поведения, которые обеспечивают стабильность социального взаимодействия индивидов). <-...->
5. абзац (социолог, 1 курс бак)
Области человеческой деятельности: наука, образование, прикладные дисциплины <-...->
6. абзац (социолог, 1 курс бак)
Но то, что текст является публицистическим, не исключает наличия в нем терминологической части:« социальное пространство совокупность социальных отношений, процессов и позиций),« социальный процесс система социальных взаимодействий и явлений),« социология наука об общественный процессах и их взаимодействии в социологическом пространстве),« социальное взаимодействие (система социальных действий),« социальная норма общепринятые

MICUSP

MICUSP Simple BETA
Michigan Corpus of Upper-Level Student Papers

SEARCH | CLEAR SEARCH | include notes & references

You are browsing papers in 16 disciplines at 4 levels of 7 paper types with 8 textual features.

▶ STUDENT LEVELS

▶ NATIVENESS

▶ TEXTUAL FEATURES

▼ PAPER TYPES

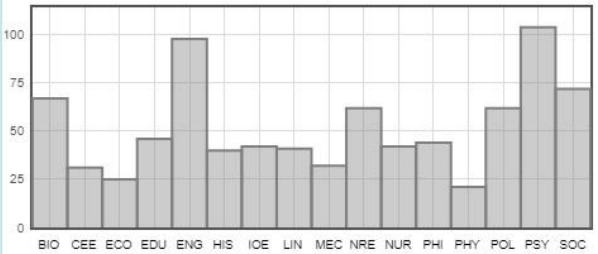
- [x] No Restriction
- [ ] Argumentative Essay
- [ ] Creative Writing
- [ ] Critique/Evaluation
- [ ] Proposal
- [ ] Report
- [ ] Research Paper
- [ ] Response Paper

▼ DISCIPLINES

- [x] No Restriction
- [ ] Biology (BIO)
- [ ] Civil & Environmental Engineering (CEE)
- [ ] Economics (ECO)
- [ ] Education (EDU)
- [ ] English (ENG)
- [ ] History & Classical Studies (HIS)
- [ ] Industrial & Operations Engineering (IOE)
- [ ] Linguistics (LIN)

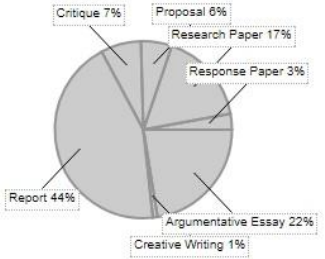
DISTRIBUTION ACROSS DISCIPLINES

CLICK TO SELECT | CLEAR SELECTION



DISTRIBUTION ACROSS PAPER TYPES

CLICK TO SELECT | CLEAR SELECTION

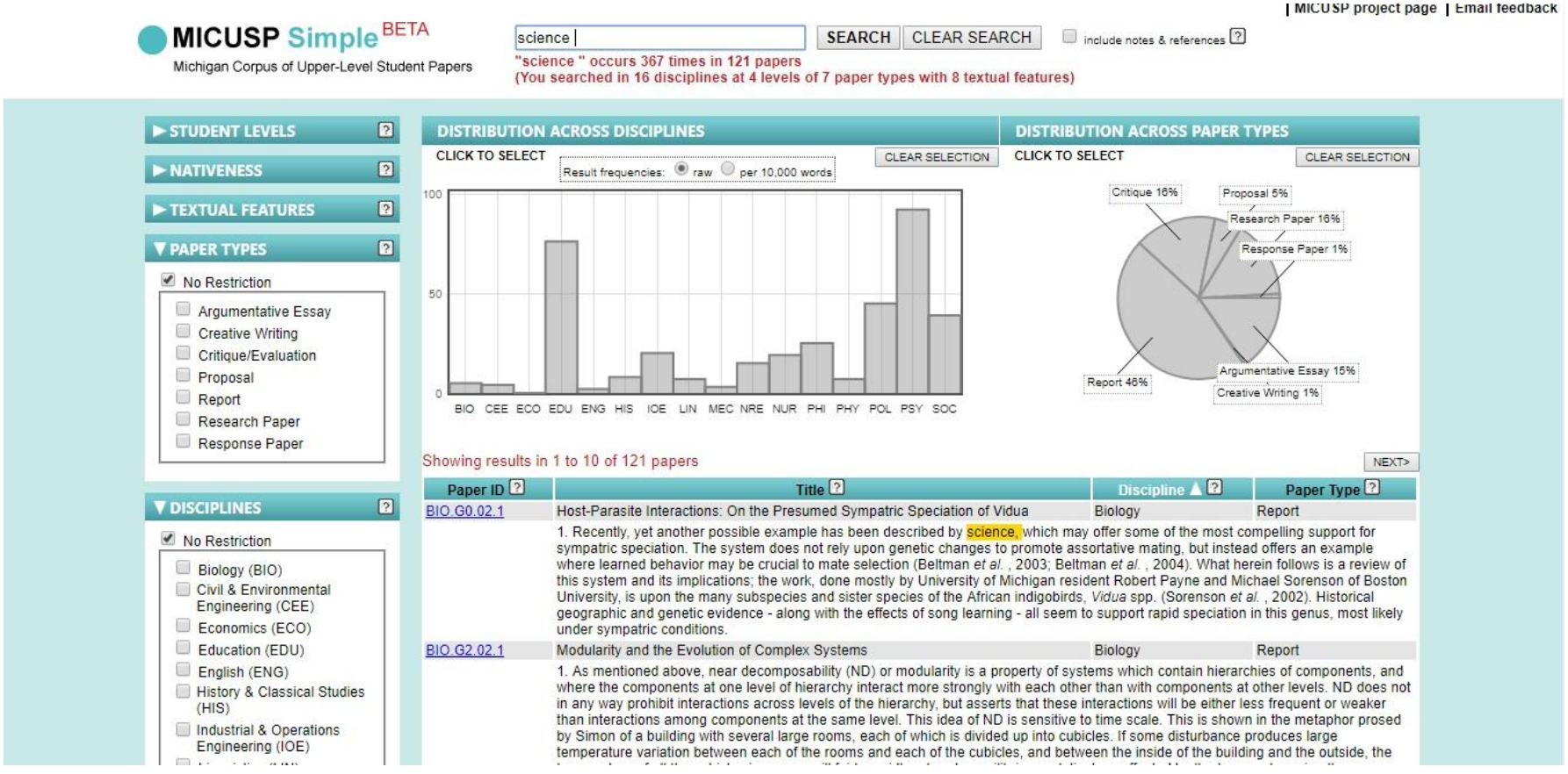


Showing 1 to 20 of 829 papers

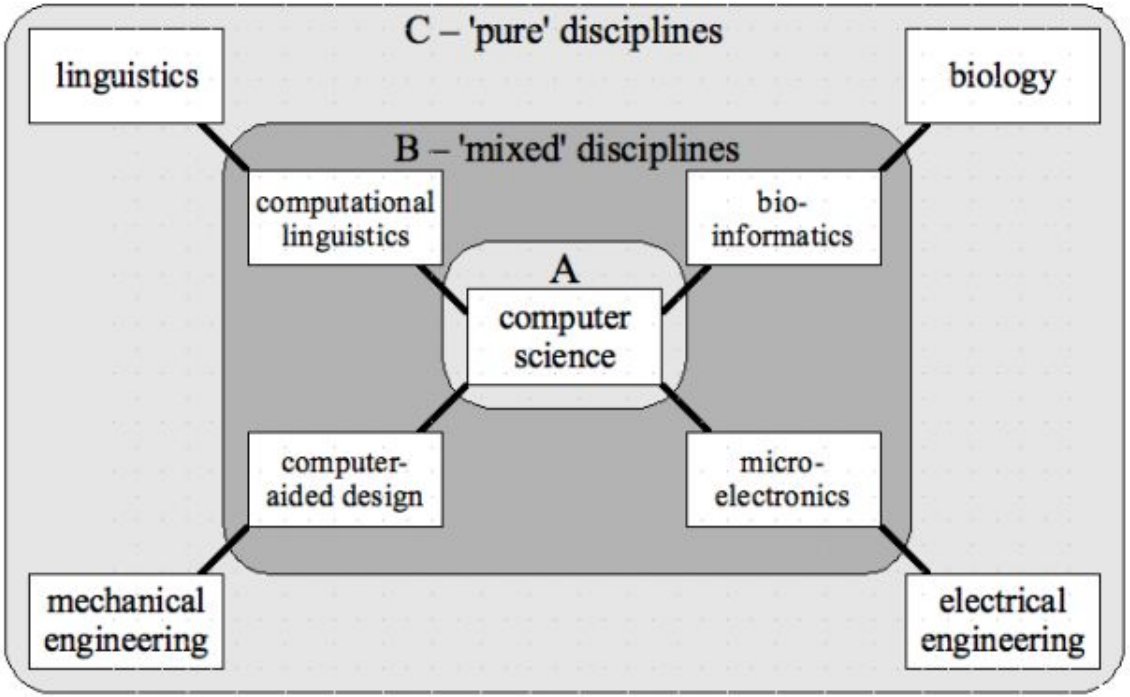
Print | Download | Link to all results | NEXT>

Paper ID	Title	Discipline	Paper Type
BIO.G0.15.1	Invading the Territory of Invasives: The Dangers of Biotic Disturbance	Biology	Argumentative Essay
BIO.G1.04.1	The Evolution of Terrestriality: A Look at the Factors that Drove Tetrapods to Move Onto Land	Biology	Argumentative Essay
BIO.G3.03.1	Intracellular Electric Field Sensing using Nano-sized Voltmeters	Biology	Argumentative Essay
BIO.G0.11.1	Exploring the Molecular Responses of Arabidopsis in Hypobaric Environments: Identifying Possible Targets for Genetic Engineering	Biology	Proposal
BIO.G1.01.1	V. Cholerae: First Steps towards a Spatially Explicit Model	Biology	Proposal
BIO.G1.07.1	Zebrafish and PGC mis-migration	Biology	Proposal
BIO.G2.06.1	A Conserved Role of Cas-Spg System in Endoderm Specification during Early Vertebrate Development	Biology	Proposal
BIO.G3.02.1	Linking scales to understand diversity	Biology	Proposal
BIO.G0.01.1	The Ecology and Epidemiology of Plague	Biology	Report
BIO.G0.02.1	Host-Parasite Interactions: On the Presumed Sympatric Speciation of Vidua	Biology	Report
BIO.G0.02.2	Sensory Drive and Speciation	Biology	Report

MICUSP


| MICUSP project page | Email feedback
MICUSP Simple BETA
Michigan Corpus of Upper-Level Student Papers
science
SEARCH
CLEAR SEARCH
include notes & references
"science " occurs 367 times in 121 papers
(You searched in 16 disciplines at 4 levels of 7 paper types with 8 textual features)
▶ STUDENT LEVELS
▶ NATIVENESS
▶ TEXTUAL FEATURES
▼ PAPER TYPES
No Restriction
Argumentative Essay
Creative Writing
Critique/Evaluation
Proposal
Report
Research Paper
Response Paper
▼ DISCIPLINES
No Restriction
Biology (BIO)
Civil & Environmental Engineering (CEE)
Economics (ECO)
Education (EDU)
English (ENG)
History & Classical Studies (HIS)
Industrial & Operations Engineering (IOE)
DISTRIBUTION ACROSS DISCIPLINES
CLICK TO SELECT
CLEAR SELECTION
Result frequencies: raw per 10,000 words
100
50
0
BIO CEE ECO EDU ENG HIS IOE LIN MEC NRE NUR PHI PHY POL PSY SOC
DISTRIBUTION ACROSS PAPER TYPES
CLICK TO SELECT
CLEAR SELECTION
Critique 16%
Proposal 5%
Research Paper 16%
Response Paper 1%
Argumentative Essay 15%
Creative Writing 1%
Report 46%
Showing results in 1 to 10 of 121 papers
NEXT>
Paper ID
Title
Discipline
Paper Type
BIO.G0.02.1
Host-Parasite Interactions: On the Presumed Sympatric Speciation of Vidua
Biology
Report
1. Recently, yet another possible example has been described by science, which may offer some of the most compelling support for sympatric speciation. The system does not rely upon genetic changes to promote assortative mating, but instead offers an example where learned behavior may be crucial to mate selection (Beltman et al. , 2003; Beltman et al. , 2004). What herein follows is a review of this system and its implications; the work, done mostly by University of Michigan resident Robert Payne and Michael Sorenson of Boston University, is upon the many subspecies and sister species of the African indigobirds, Vidua spp. (Sorenson et al. , 2002). Historical geographic and genetic evidence - along with the effects of song learning - all seem to support rapid speciation in this genus, most likely under sympatric conditions.
BIO.G2.02.1
Modularity and the Evolution of Complex Systems
Biology
Report
1. As mentioned above, near decomposability (ND) or modularity is a property of systems which contain hierarchies of components, and where the components at one level of hierarchy interact more strongly with each other than with components at other levels. ND does not in any way prohibit interactions across levels of the hierarchy, but asserts that these interactions will be either less frequent or weaker than interactions among components at the same level. This idea of ND is sensitive to time scale. This is shown in the metaphor prosed by Simon of a building with several large rooms, each of which is divided up into cubicles. If some disturbance produces large temperature variation between each of the rooms and each of the cubicles, and between the inside of the building and the outside, the

Дармштадтский корпус научных текстов (DaSciTex)



DaSciTex

A - Computer science	Journal of Algorithms Journal of Computer and System Science	2004-2006 2005-2007	2,353,328
B1 - Computational linguistics	J. of Computational Linguistics Machine Translation J. of Natural Language Engineering	2003-2006 1998-2004 2006 (12:1)	1,295,652
B2 - Bioinformatics	Bioinformatics J. of Computational Biology	2006 2005 (12)	2,016,488
B3 - Computer-aided design	Advanced Engineering Informatics Advances in Engineering Software	2004-2006 2005-2007	1,545,433
B4 - Microelectronics	IEEE Transactions on VLSI Journal on Embedded Systems	2005-2006 2006 (1)	1,543,759
C1 - Linguistic	Language J. of Linguistics Functions of Language Linguistic Inquiry	2003-2006 2006 (42:1) 2005-2006 2005-2006	1,741,711
C2 - Biology	Gene Nucleic Acid Research	2004-2006 2006	2,329,493
C3 - Mechanical engineering	Chemical Engineering and Processing Chemical Engineering Science International J. of Heat and Mass Transfer	2006-2007 2006(10)-(1)2007 2006(10)-(1)2007	1,994,923
C4 - Electrical engineering	Mechatronics Automatica Control Engineering Practice	2006 2006-2007(1) 2006(7-12)-2007(1-3)	1,663,657
Σ			16,484,444

Проблема определения нормы

“Диапазон нарушений колеблется от грубых ошибок до вполне нормальных вариантов. Все эти случаи объединяет одно: они вызвали негативную реакцию читателя, оказались для него неприемлемыми или просто менее привычными, неудобными, нестандартными” (Кукушкина).

“Подавляющая часть норм, которые организуют речемыслительную деятельность, носит «подсознательный» характер” (Кукушкина).

“Правила могут „управлять” речевым поведением и даже способностью критически оценивать предложения таким образом, что человек не отдает себе отчета в их существовании” (Стронсон).

Три общих параметра оценки качества продукта речевой деятельности:

1) эффективность, ясность, понятность, простота и т. п.;

2) адекватность, точность;

3) корректность, приемлемость.

НО - оценка правильности все равно остается субъективной

1) носитель языка не всегда способен отличить правильно построенное высказывание от неправильного;

2) существуют значительные расхождения в оценках правильности;

3) реакции говорящих не дают ясного представления о том, какие из факторов — семантические или синтаксические — являются источником неприемлемости конкретного предложения.

Классификация ошибок

НА УРОВНЕ СЛОВА: орфографические, словообразовательные, грамматические, лексические

НА УРОВНЕ СЛОВОСОЧЕТАНИЯ

НА УРОВНЕ ПРЕДЛОЖЕНИЯ: синтаксические, коммуникативные

НА УРОВНЕ ТЕКСТА: логические нарушения, грамматические нарушения, информационно-коммуникативные нарушения

О.В. Кукушкина: Основные типы речевых неудач в русских письменных текстах.

1) неудачное структурирование информации;

2) неудачная квалификация общего типа логических отношений, связывающих основные элементы пропозиции («предмет» и «признак»);

3) неудачная квалификация положения дел, описываемого пропозитивной структурой;

4) неудачная категориальная квалификация отдельных элементов положения дел, описываемого пропозицией;

5) неудачная актуализация категорий, использованных для описания элементов положения дел;

6) неудачный выбор означающего;

7) неудачная запись означающего.

Автоматическая оценка качества научного текста – как всё начиналось

- Первые работы, связанные с автоматической оценкой качества текста, принадлежат Ellis B. Page (The Imminence of... Grading Essays by Computer, 1966);
- Работа над системой автоматической оценки эссе началась в 1964 г. и продолжалась вплоть до 90-х годов

Автоматическая оценка качества научного текста – как всё начиналось

- Множество формальных критериев: наличие заголовка, средняя длина предложения, количество апострофов, количество двоеточий…
- Оценки, выставленные системой, во многом совпадали с человеческими

Не совсем автоматическая оценка качества научного текста

- Пример критериев оценивания вручную: Bates College Peer Review Form
- Оценивается работа целиком;
- Критерии включают ясность постановки цели, качество данных и т.п., некоторые из них довольно субъективны;
- Есть также формальные критерии – стиль, отсутствие грамматических ошибок…

Автоматическая оценка качества научного текста - SWAN

- SWAN – Scientific Writing AssistaNt (Kinnunen et al., 2012)
- Система, помогающая сделать научный текст reader-friendly;
- Фокусируется на том, что формирует первое впечатление о работе;
- Оценивается не общее качество, а связность и плавность изложения.

Автоматическая оценка качества научного текста - SWAN

- Внимание на:
  - Заголовок;
  - Абстракт;
  - Введение;
  - Заключение.
- Используются автоматические (напр., поиск пассивного залога) и ручные (напр., выделение значимых ключевых слов) методы оценивания;
- Внимание на связь между фрагментами.

Автоматическая оценка качества научного текста - Леменков

- Автоматическая оценка разделов научных текстов – Д.Д. Леменков, 2016
- Оценивается заголовок и аннотация;
- Сравнивается результат работы алгоритма, обученного на данных от экспертов, и алгоритма, использующего правила;
- Достигнута достаточно высокая согласованность экспертных и автоматически выставленных оценок.

Автоматическая оценка качества научного текста - Леменков

- Оценка качества заголовка:
  - Понятность;
  - Возможность поиска статьи по заголовку;
  - Привлекательность заголовка.
- Оценка качества аннотации:
  - Согласованность с заголовком;
  - Привлекательность;
  - Точность.


HD
главный
КОТИК
страны

Швец А.В. - Взаимодействие информационных и лингвистических методов в задачах анализа качества научных текстов (диссертация).

Оценка качества:

-оценка лексики и синтактико-семантических структур текста

-оценка наличия лингвистических ошибок

-оценка наличия псевдонаучных фрагментов

-оценка формальной структуры текста, т. е. наличия в тексте необходимых разделов (например, описания результатов).

Рисунок 1 – Признаки, характеризующие качество текстов научной сферы

Примеры ошибок

Пример 1 (нарушение требований к лексике): «И что об этом думают сами языковеды? Не стану добавлять имеющуюся словесную чепуху с целью придания наукообразия ссылками на разнообразные мнения на сей счет. Их без труда можно найти в Интернете».

Пример 2 (нарушение правил согласования): «Такие факторы как возраст, образование, социальный статус обычно оказывает существенное влияние на речевое поведение носителя языка».

Пример 3 (нарушение семантической связности): «Сформулировать и доказать о свойствах прямоугольных треугольников».

Пример 4 (лексическая избыточность): «То, что я назвал понятием, в этих школах обычно называют содержанием понятия, хотя содержание этого содержания может несколько варьироваться от школы к школе и соответственно отличаться от моего».

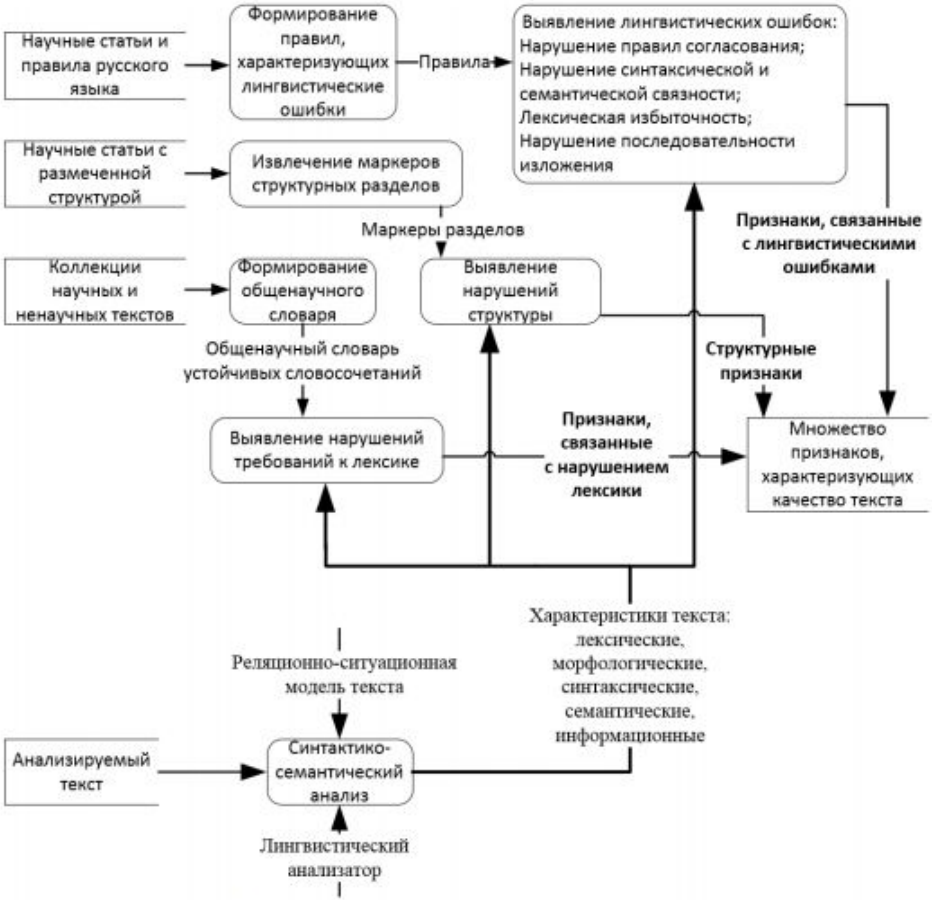


Рисунок 2 – Предлагаемая в работе схема выявления признаков, характеризующих качество текстов научной сферы

Признаки для определения псевдонаучности

-слова (например, "торсионный", "гармонизировать", "чрезвычайно", "неправота");

-словосочетания с синтаксическими и семантическими связями (например, "повсеместное наличие", "необъяснимая аномалия", "усматривать в модели", "убедительно показать", "память воды");

-обобщения словосочетаний (например, "память <сущ.>", "<прил.> аномалия", "усматривать в <сущ.>");

-триграммы (например, "я якобы сразу", "и почти нигде", "совершенно очевидно то", "сейчас наукой доказано").