

3. Фомичев, А. А. Оценка дисциплинарных компетенций студентов / А. А. Фомичев, З. Ю. Филиппова // Известия ТулГУ. Серия: Технические науки. – 2016. – Вып. 11. – Ч. 1. – С. 273–281.

УДК [378.012:378.062]:004.4

**Якунин Ю. Ю., Даничев А. А.**

**АНАЛИЗ НЕСТРУКТУРИРОВАННЫХ ДАННЫХ ОБРАТНОЙ СВЯЗИ  
В ПЕРСОНАЛЬНОЙ ОБРАЗОВАТЕЛЬНОЙ СРЕДЕ**

*Юрий Юрьевич Якунин*

*кандидат технических наук, доцент*

*yakuninyu@mail.ru*

*Алексей Александрович Даничев*

*кандидат технических наук*

*aleksey-d@inbox.ru*

*ФГАОУ ВО «Сибирский федеральный университет», Россия, Красноярск*

**ANALYSIS OF UNSTRUCTURED DATA OF FEEDBACK IN PERSONAL  
LEARNING ENVIRONMENT**

*Yury Yurievich Yakunin*

*Aleksey Aleksandrovich Danichev*

*Siberian Federal University, Russia, Krasnoyarsk*

*Аннотация. В статье рассмотрен анализ текстовых ответов анкетирования студентов в персональной образовательной среде. Раскрываются возможности анализа текста инструментами InterSystems iKnow. Изложены основные аспекты автоматического выявления эмоциональных оттенков отзывов.*

***Abstract.** In the article the analysis of text responses from the survey of students in a personal learning environment. The possibilities of text analysis by InterSystems iKnow tools are revealed. The main aspects of automatic detection of emotional shades of feedback are described.*

***Ключевые слова:** персональная образовательная среда, семантический анализ, тональность текстов.*

***Keywords:** personal learning environment, semantic analysis, sentiment analysis of text.*

Последние несколько лет в Сибирском федеральном университете эксплуатируются и модернизируются инструменты образовательной среды студентов [16]. В Институте космических и информационных технологий (ИКИТ) на базе личного кабинета студента реализован сервис анкетирования студентов, который позволяет получать обратную связь в структурированном и текстовом видах по различным аспектам образовательного процесса.

Анкетирование студентов проводится регулярно после каждого семестра и включает перечень вопросов, касающихся:

- отношения к работе выпускающей кафедры;
- работы учебно-организационного отдела (деканата);
- характеристики преподавания предметов и обеспечивающих их материалов.

Ответы возможны как в виде оценке по бальной шкале, так и в свободной форме в виде текста. В ИКИТ автоматизированное анкетирование проводится с 2015 года и к настоящему времени насчитывает в общей сложности 130 тыс. ответов от 1200 студентов [14].

Основным направлением обработки результатов данного анкетирования является анализ бальных оценок, их сопоставление с результатами итоговой аттестации студентов [9, 11–13, 15]. При таком подходе ответы в свободной форме используются крайне ограниченно. Например, если некоторая группа

студентов поставила аномально низкие баллы лектору, имеет смысл прочитать их отзывы и выяснить причину.

Анализ текста ответов прежде всего позволяет выявить скрытые показатели, на которые студенты чаще всего обращают внимание. Такие показатели будут актуальны, понятны, интересны самим студентам. Новые показатели должны повысить качество обратной связи при анкетировании.

Еще одно перспективное направление исследований — автоматическое выявление эмоциональных оттенков отзывов. Этот инструмент так же поможет использовать новые показатели и повысить эффективность проверки достоверности бальных оценок.

Несмотря на то, что до 95% всех данных в мире [7] представлено в неструктурированной форме, чаще всего такие данные не используются, так как их анализ сложен, длителен и дорог. Автоматический анализ тональности текстов, т. е. выявление мнения автора текста по поводу предмета, обсуждаемого в тексте, является одной из активно развиваемых технологий в сфере автоматической обработки текстов в последнее десятилетие [8]. Большое количество работ посвящено анализу тональности отзывов пользователей, которые они оставляют в рекомендательных сервисах [5]. Как правило, решаются задачи анализа тональности по отношению к заданным сущностям, упомянутым в тексте [1]. Более детальным уровнем является анализ мнения по конкретным свойствам или частям (аспектам) сущности, по которым автор текста может высказывать разную тональность мнения.

Обычно предполагается, что тональность выражается с помощью оценочной лексики, что представляет собой эксплицитный (явный) способ выражения оценок. Значительно усложняют задачу имплицитные (скрытые) оценки, с использованием оценочных фактов [5] или слов с коннотациями [3], а так же сарказм, ирония. Основные словари оценочной лексики для русского языка: RuСентиЛекс [10], ProductSentiRus [2], Linis-crowd [6]. В сфере образования таких словарей для русского языка авторам неизвестно. В нашем случае

в распоряжении имеется одновременно наличие отзывов о предметах и преподавателях как в текстовой форме, так и в виде численной оценки. Это позволяет частично выявить положительные или отрицательные коннотации слов.

Полная автоматизация выявления оценочных слов вряд ли возможна. Внутри предметной области оценочные слова могут нести положительную или отрицательную тональность в зависимости от аспекта объекта, к которому они применяются. Например, слово «долго» может быть как отрицательным, так и положительным. Если говорят, что «задания долго проверяют», то это плохо. Если говорят, что «материалы лекций долго помнишь», то это хорошо.

В данной работе семантический анализ текста, выполнен инструментом InterSystems iKnow [4]. Технология iKnow позволяет выявлять ключевые понятия и связи в неструктурированных данных. В отличие от традиционных технологий семантического анализа iKnow избавляет разработчиков от необходимости предварительного создания словарей терминов или фраз — она находит их самостоятельно. На рисунке 1 представлен результат анализа iKnow.

Также **я** хочу предложить ввести больше преподаватель для сдачи лабораторных работ.  
У многих работы были выполнены, но **очереди на сдачу** **не** дождёшься  
**Ладе** вообще **солнышко**, а **не** **преподаватель**, побольше бы **таких**

**Желтый:** понятие.

Подчеркнуто: отношение.

*Курсив:* не важные фразы.

**Красный:** фраза атрибута отрицания.

Рисунок 1 — Семантический анализ предложений

Семантическим анализом около ста тысяч текстовых отзывов студентов было выделено 1500 информативных отзывов, содержащих 100 основных понятий из предметной области (общее количество понятий более 10 тысяч). Так как различные словоформы одного слова считаются различными понятиями, то необходимо определиться с основными понятиями, и сформировать словари синонимов и словоформ. Например, словарь «преподаватель» содержит слова: преподаватели; лектору; ассистента и пр. В таблице 1 представлены

наиболее часто встречающиеся понятия — «преподаватель», «студент», «лекции» и «предмет».

Таблица 1 — Основные понятия из отзывов

<b>Понятие</b>	<b>Частота</b>
преподаватель	131
студент	111
лекции	108
предмет	90

Следующий этап анализа — определение понятий и отношений характерных для положительной или отрицательной тональности в отзывах.

Инструменты iKnow позволяют понять, какие понятия наиболее тесно связаны с искомыми терминами (таблицы 2, 3), и какие отношения задают эту связь (таблица 4).

Таблица 2 — Похожие понятия для «преподаватель»

<b>№</b>	<b>Понятие</b>	<b>Частота</b>
1	«преподаватель»	131
2	лучший «преподаватель»	4
3	прекрасный «преподаватель»	4
4	компетентность «преподаватель»	3
5	смена «преподаватель»	3
6	компетентность некоторых «преподаватель»	2
...	...	...
78	хороший «преподаватель»	1

Таблица 3 — Связанные понятия для «преподаватель»

<b>№</b>	<b>Понятие</b>	<b>Частота</b>
1	предмет	7
2	лекции	4
3	весь семестр	3
4	лабораторные	3
5	вопросы	3
6	ощущение	2
...	...	...
93	слышимость очень низкая	1

Таблица 4 — Пары «отношение и понятие» для «преподаватель»

№	Сущности (понятие <u>отношение</u> понятие)	Частота
1	<u>если</u> преподаватель	3
2	<u>потому что</u> преподаватель	3
3	<u>сменить</u> преподаватель	3
4	<u>заменить</u> преподаватель	2
5	<u>смените</u> преподаватель	2
6	<u>предмет потому что</u> преподаватель	2
7	преподаватель <u>дает</u> лабораторные	2
...	...	...
163	преподаватель <u>не использовал слышимость</u> очень низкая	1

В результате анализа были сформированы словари с наиболее часто выраженными оценками (положительной или отрицательной тональности) для главных понятий. Так, для понятия «преподаватель» выявлены оценки: «хороший»; «плохой»; «строгий»; «требовательный». Словари оценочной лексики можно представить как соответствие оценок и понятий (таблица 5).

Завершающим развитием словарей оценочной лексики будет сопоставление понятий и отношений с бальными оценками на те же вопросы, что позволит вычислить вероятности их оценочности или весовые коэффициенты [8].

Таблица 5 — Сопоставление понятий с оценками тональности

№	Понятие из текста	Термин словаря
1	лучший преподаватель	хороший преподаватель
2	отличный преподаватель	хороший преподаватель
3	очень строгий преподаватель	строгий преподаватель
4	понимающий преподаватель	хороший преподаватель
5	прекрасный преподаватель	хороший преподаватель
6	преподаватель очень толковый	хороший преподаватель
7	самый лучший преподаватель	хороший преподаватель
8	Преподаватель слишком требовательный	требовательный преподаватель
...	...	...
83	одна из худших дисциплин	плохой предмет

В итоге ручной обработки результатов семантического анализа:

- выделены 4 основных понятия, раскрываемых в отзывах студентов;
- для каждого понятия сформированы оценки с положительной или отрицательной тональностью;

- для большинства автоматически выделенных значимых понятий задано соответствие с основными понятиями.

Для объединения результатов семантического анализа и всех имеющихся данных (структурированных и неструктурированных) применяется аналитическая технология InterSystems DeepSee. Оценка эмоциональной тональности отзывов применена для проверки достоверности отзывов студентов. Новые показатели позволили выявлять неординарные ситуации, например негативное отношение к преподавателю, независимо от успеваемости студентов.

Другой важный результат — это внесение изменений в анкету на основании результатов анализа ответов студентов. Изменение в анкету вносятся после подведения итогов каждый семестр. Одним из дополнений стал вопрос об отношении преподавателя к студентам.

Таким образом, выявленные основные понятия и их выраженные оценки формируют новые показатели, улучшающие качество обратной связи при анкетировании, а накопленные результаты обработки отзывов студентов являются базой для реализации автоматического анализа тональности текстов в сфере контроля качества образования.

### *Список литературы*

1. Overview of replab 2012: Evaluating online reputation management systems / Amigo E., Corujo A., Gonzalo J., Meij E., Rijke Md. // Proceedings of the CLEF 2012 Labs and Workshop Notebook Papers. – 2012. – P. 1–24.
2. Chetviorkin, I. Extraction of Russian Sentiment Lexicon for Product Meta-Domain / Chetviorkin I., Loukachevitch N. // Proceedings of COLING-2012. – 2012. – P. 593–610.
3. Connotation Lexicon: A Dash of Sentiment Beneath the Surface Meaning / Feng S., Kang J. S., Kuznetsova P., Choi Y. // Proceedings of ACL. – 2013. – P. 1774–1784.
4. InterSystems Documentation [Electronic resource]. – URL: <https://docs.intersystems.com> (дата обращения: 21.01.2019).

5. Liu, B. Sentiment analysis and Subjectivity / Liu B. // Handbook of Natural Language Processing. – CRC Press, Taylor and Francis Group, Boca Raton, 2010. – P. 1–38.

6. Алексеева, С. В. Linis-crowd. org: лексический ресурс для анализа тотальности социально-политических текстов на русском языке / С. В. Алексеева, С. Н. Кольцов, О. Ю. Кольцова // Компьютерная лингвистика и вычислительные онтологии : труды XVIII объединенной конференции. Интернет и современное общество (IMS–2015). – Санкт-Петербург, 2015. – С. 25–34.

7. Встраиваемые технологии [Электронный ресурс]. – Режим доступа: <http://www.intersystems.com/ru/our-products/embedded-technologies/iknow/> (дата обращения: 21.01.2019).

8. Автоматическая обработка текстов на естественном языке и анализ данных : учебное пособие / Н. В. Лукашевич [и др. ]. – Москва : Издательство НИУ ВШЭ, 2017. – 269 с.

9. Медведев, А. В. О математическом моделировании образовательного процесса в университете / А. В. Медведев, Ю. Ю. Якунин, Д. И. Ярещенко // Высшее образование сегодня. – 2016. – № 11. – С. 45–51.

10. Словарь оценочных слов и выражений русского языка RuСентиЛекс [Электронный ресурс]. – Режим доступа: <http://www.labinform.ru/pub/rusentilex> (дата обращения: 21.01.2019).

11. Фомина, Е. Е. Подготовка и анализ результатов анкетирования с применением математических методов статистики / Е. Е. Фомина // Социосфера. – 2018. – № 2. – С. 194–198.

12. Фомина, Е. Е. Факторный анализ и категориальный метод главных компонент: сравнительный анализ и практическое применение для обработки результатов анкетирования статистики / Е. Е. Фомина // Гуманитарный вестник. – 2017. – № 10 (60). – С. 3.

13. Фомина, Е. Е. Методика обработки результатов анкетирования с использованием методов многомерной и параметрической статистики / Е. Е. Фомина, Н. К. Жиганов // Вестник Пермского национального исследовательского



политехнического университета. Социально-экономические науки. – 2017. – № 1. – С. 106–115.

14. Погребников, А. К. Анализ обратной связи в персональной образовательной среде / А. К. Погребников, Ю. Ю. Якунин // Информатика и образование. – Москва : Образование и Информатика, 2018. – № 10 (299). – С. 36–41.

15. Медведев, А. В. О математическом моделировании образовательного процесса в университете / А. В. Медведев, Ю. Ю. Якунин, Д. И. Ярещенко // Высшее образование сегодня. – Москва : ЛОГОС, 2016. – Вып. 11. – С. 45–51.

16. Якунин, Ю. Ю. Персональная образовательная среда в системе управления институтом / Ю. Ю. Якунин, А. К. Погребников // Информатика и образование. – Москва : Образование и Информатика, 2017. – Вып. 2 (281). – С. 50–55.