

## Программа формирования контекста для электронной коллекции «Тувинские героические сказания»

*Ч.М. Монгуш*

*Тувинский государственный университет, Кызыл*

**Аннотация:** Электронная коллекция «Тувинские героические сказания» содержит оцифрованные тексты произведений, их метаописания, справочники сказителей и языковых стандартов. Такая информация хорошо структурирована и допускает представление в виде таблицы «объект-признак», называемой контекстом, который отражает наличие или отсутствие признаков, характерных для исследуемых произведений. Имея в наличии контекст, с помощью математических методов, можно решать различные задачи машинного обучения применительно к текстам тувинского героического эпоса. В статье предлагается компьютерная программа FCAContext формирования контекстов для электронной коллекции «Тувинские героические сказания». Подробно описаны основные возможности и функции данной программы. Данное исследование проводилось при поддержке Российского гуманитарного научного фонда, грант 16-34-1-01033.

**Ключевые слова:** электронная коллекция документов, тувинский героический эпос, бинарный контекст.

Над созданием электронных коллекций произведений тувинского фольклора ученые Тувинского государственного университета и научно-исследовательского центра «Тюркология» работают не одно десятилетие [1]. Сформированный фонд рукописных и магнитофонных записей всех жанров тувинского фольклора находится в научном архиве Тувинского института гуманитарных и прикладных социально-экономических исследований. В архиве института хранятся около 300 записей эпических произведений [2]. Однако свет увидели немногие, поскольку данные произведения хранятся в рукописном виде или напечатаны в старых ветхих изданиях. В настоящее время имеются лишь 14 напечатанных собраний тувинских героических произведений [3]. На основе этих публикаций под руководством профессора Тувинского государственного университета М.В. Бавуу-Сюрюн создана электронная коллекция «Тувинские героические сказания» [4]. В коллекции «Тувинские героические сказания» хранятся не только оцифрованные тексты

---

произведений, но и их метаописания, включая сведения о сказителях. Такая информация, как правило, хорошо структурирована и допускает представление в виде таблицы «объект-признак», называемой контекстом. Контекст отражает наличие или отсутствие признаков, характерных для исследуемых произведений. Имея в наличии контекст, с помощью математических методов можно решать различные задачи машинного обучения [5–9].

В статье предлагается компьютерная программа FCAContext для формирования контекста по электронной коллекции «Тувинские героические сказания». Подробно описаны основные функции и реляционные таблицы, используемые в программе и формируемые данной программой.

### **Контекст как модель описания коллекции произведений**

Под контекстом понимается тройка  $K = (G, M, I)$ , где  $G$  и  $M$  – два непустых конечных множества объектов и признаков исследуемой предметной области соответственно,  $I \subseteq G \times M$  – отношение инцидентности между множествами  $G$  и  $M$ . Отношение  $I$  интерпретируется следующим образом: пара  $(g, m)$ ,  $g \in G$  и  $m \in M$ , означает, что объект  $g$  имеет признак  $m$  и наоборот, признак  $m$  присущ объекту  $g$  [10, 11].

Для формирования контекста необходимо выбрать состав объектов и признаков, исходя из поставленной задачи исследования. Например, для решения задачи выявления авторского стиля сказителей по описанию снаряжения коня в тувинских героических сказаниях в качестве исходных данных можно использовать контекст, представленный в таблице № 1. В таблице № 1 объектами контекста являются четыре произведения:  $G = \{\text{Демир-Шилги аъттыг Тевене-Мөге, Мөге Шагаан-Тоолай, Танаа-Херел, Каңгывай-Мерген}\}$ ; а признаками – языковые стандарты, характеризующие элемент снаряжения коня богатыря:  $M = \{\text{седло, хлыст, потник}\}$ .

---

Таблица № 1

Контекст  $K$  анализируемых произведений тувинского героического эпоса

№	Название эпоса	Языковые стандарты, характеризующие элементы снаряжения коня			Сказитель
		седло	хлыст	потник	
1	Демир-Шилги аьттыг Тевене-Мөге	Арт болган алчайган-калчайган эзер		Шөл болган шөйүлген чаттылган чонок	Ооржак Чанчы-Хөө Чапаажыкович
2	Мөге Шагаан-Тоолай	Арт болган аңгайган-каңгайган кызыл чуңгуу эзер	Алдын допурзак кымчы	Хөл болган хөлбең кара чонок	Ондар Тевек-Кежеге
3	Танаа-Херел	Арт болган аңгайган-каңгайган кызыл чуңгуу эзер		Хөл болган хөлбең кара чонок	Ондар Тевек-Кежеге
4	Каңгывай-Мерген	Арт болган алдын каңгай эзер	Алдын допурзак кымчы		Тюлюш Баазаңай Халдаевич

Большинство математических методов, которые решают данную задачу, работают с бинарными контекстами [5, 11]. Поэтому возникает необходимость приведение контекста  $K = (G, M, I)$  к бинарному виду. Для этого выполняется шкалирование качественных признаков в контексте, т. е. устанавливается соответствие между текстовым значением признака и его эквивалентом. Наиболее распространенным приемом шкалирования служит использование номинальной шкалы. Примеры такого шкалирования дают таблицы № 2 и № 3. В таблице № 2 названия произведений заменены их порядковыми номерами. Единичный (нулевой или пустой) элемент этой таблицы указывает на то, что соответствующее литературное произведение обладает (не обладает) тем или иным признаком. Значение признаков заменены соответствующими идентификаторами, которые представлены таблице № 3.

Таблица № 2

Бинарный контекст

№	$m_1$	$m_2$	$m_3$	$m_4$	$m_5$	$m_6$	$m_7$	$m_8$	$m_9$
1	1				1		1		
2		1		1		1		1	
3		1		1		1		1	
4			1						1

Таблица № 3

Список языковых стандартов

Идентификатор признака	Значение признака
$m_1$	Арт болган алчайган-калчайган эзер
$m_2$	Арт болган аңгайган-каңгайган кызыл чуңгуу эзер
$m_3$	Арт болган алдын каңгай эзер
$m_4$	Алдын допурзак кымчы
$m_5$	Шөл болган шөйүлген чаттылган чонак
$m_6$	Хөл болган хөлбең кара чонак
$m_7$	Ооржак Чанчы-Хөө Чапаажыкович
$m_8$	Ондар Тевек-Кежеге
$m_9$	Тюлюш Баазанай Халдаевич

Часто требуется создавать контексты с большим количеством объектов и признаков. Поэтому для формирования больших контекстов необходимы соответствующие информационные и программные средства.

**Описание программы**

Программа FCASContext предназначена для формирования контекстов для электронной коллекции «Тувинские героические сказания» и создана на языке программирования C# в интегрированной среде разработки Microsoft Visual Studio Community 2017. Для эксплуатации программы требуется персональный компьютер типа IBM PC Pentium IV с операционной системой Windows 7.0/8.0/10.0 и оперативной памятью от 4 Гб.

Программа FSAContext ориентирована на пользователей корпуса тувинского языка, а именно на филологов и лингвистов. Исходными данными программы являются базы данных, содержащие тексты произведений тувинского героического эпоса, справочники сказителей и языковых стандартов (рис. 1) [12]. Результатом выполнения программы является бинарный контекст, который сохраняется в текстовом файле.

Программа для работы с базами данных тувинского героического эпоса и формирования контекстов

БД "Тувинские героические эпосы" БД "Тувинские сказители" БД "Клише" Контекст

База данных тувинских героических эпосов

Номер записи	Название произведения	Год записи	Место записи	Кем записан эпос	Источн
1	Эрелзей-Мерген, Харагалзай-Мерген алыш...			К. Дамдин	Тыва тоолдар -
2	Демир-Шилги аъттыг Тевене-Меге		Бай-Тайга	Саая Санмырович Самыя	
3	Алдын Кургулдай		Эрзин	К. Тоюң	
4	Карыш-кулаш хайлыг Калчан-Шилги аътт...		Сүт-Хөл, с. Суг-Аксы	Самозапись	
5	Хартыга-Бора аъттыг Чаңгыс-Карыш		Тожу		
6	Ээр-Сарыг аъттыг Экер-оол		Сүт-Хөл, с. Суг-Аксы	Самозапись	
7	Меге Шагаан-Тоолай	1944	Сүт-Хөл, с. Алдан-Маадыр	А.М. Кызыл-оол	Тыва маадырлы
8	Танаа-Херел	1944	Сүт-Хөл, с. Алдан-Маадыр	А.М. Кызыл-оол	Тыва маадырлы
9	Тоң-Аралчын хаан	1946	Чөөн-Хемчик, с. Хөндергей	М. Идам-Суруң	
10	Каңгывай-Мерген	1946	Улуг-Хем, с. Торгалыг		
11	Баян-Тоолай		Бай-Тайга, с. Кызыл-Даг	М. Ляндуп	

Рис. 1. – Электронная коллекция «Тувинские героические сказания»

Программа FSAContext допускает редактирование ранее построенного контекста. Кроме этого, возможен просмотр текстов анализируемых произведений коллекции «Тувинские героические сказания». В программе существует возможность частично-автоматизированного поиска языковых стандартов в текстах тувинского героического эпоса. Для этого программа выявляет в тексте отдельного произведения словосочетания, которые схожи с определенным языковым стандартом, и выдает их в отдельном окне программы (рис. 2). Эксперт филолог или лингвист из данного списка выбирает релевантные стандарты, которые затем сохраняются в справочнике языковых стандартов.

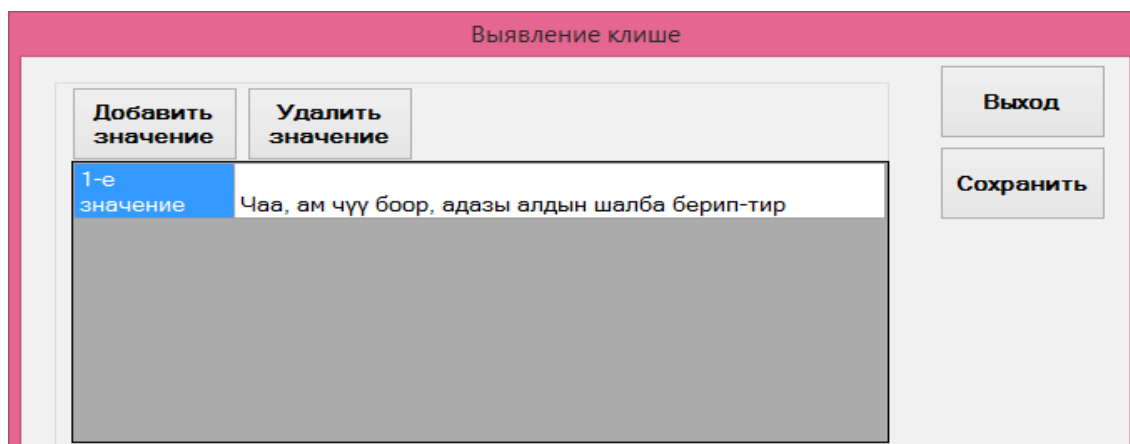


Рис. 2. – Окно программы для выявления языкового стандарта

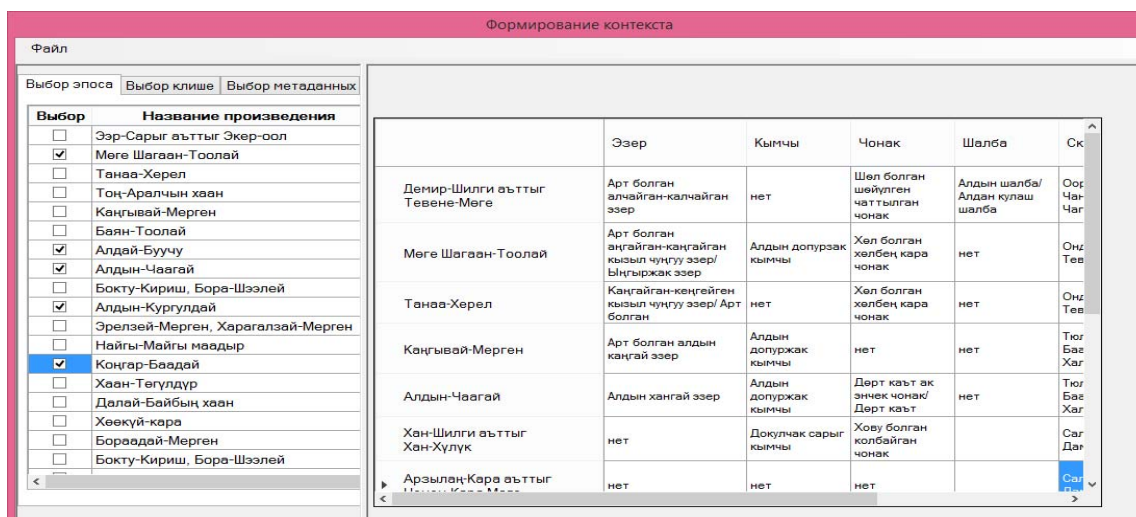


Рис. 3. – Интерфейс для формирования контекста

В программе FCASContext имеется специальный интерфейс для формирования контекстов (рис. 3). Из электронной коллекции «Тувинские героические сказания» эксперт в зависимости от поставленной задачи выбирает исследуемые произведения, отбирает сказителей и языковые стандарты из справочников. Программа FCASContext позволяет также шкалировать качественные признаки. На рис. 4 показан пример приведения исходного контекста к бинарному виду.

Формирование контекста

	Алдан кулаш шалба	Алдан кулаш алдан шалба	Оржак Чаны-Хай Чапажыкович	Ондар Телок-Кежиге	Толош Балзаяр Халдаевич	Салмак Данде Дамынович	Монгуш Хуртугол Сазыг-Хунаевич	Барун-Умникоос рн	Сут-Хо
	m26	m27	m28	m29	m30	m31	m32	m33	m34
Демир-Шиги аъттыг Тевене-Миге	1		1					1	
Миге Шалаан-Тоолал				1					1
Танаа-Херел				1					1
Кагывай-Мерген					1				
Алдан-Чагай					1				
Хан-Шиги аъттыг Хан-Хитук						1			
Арвалан-Кара аъттыг Чечен-Кара Миге						1			
Арвалан-Кара аъттыг Чечен-Кара миге						1			
Арвалан-Кара аъттыг Хунан-Кара			1					1	
Сарыг-Хемчикилтен чуртаан Таван-Халан							1		1
Хан-Шиги аъттыг Хан-Куу-Мадар							1		1
Алдан-сарыг аъттыг Алан-Кара							1		1
Элестей ашак		1	1					1	

Рис. 4. – Окно программы для представления бинарного контекста

Сформированные контексты сохраняются в текстовом файле и используются в качестве входных данных при решении филологических и лингвистических задач, решаемых в корпусе тувинского языка.

*Данное исследование проводилось при поддержке Российского гуманитарного научного фонда, грант 16-34-1-01033.*

### Литература

1. Бавуу-Сюрюн М.В. Вопросы создания электронных ресурсов тувинского языка: некоторые итоги, неотложные задачи и перспективы // Новые исследования Тувы. 2016. № 4. URL: [nit.tuva.asia/nit/article/view/610](http://nit.tuva.asia/nit/article/view/610)
2. Орус-оол С.М. Тувинские героические сказания (текстология, поэтика, стиль). М.: Макс пресс, 2001. 422 с.
3. Ондар М. В. База данных текстов тувинского героического эпоса: первый этап // Новые исследования Тувы. 2016. № 4. URL: [nit.tuva.asia/nit/article/view/616](http://nit.tuva.asia/nit/article/view/616).
4. Бавуу-Сюрюн М.В., Далаа С.М., Монгуш Ч.М., Ондар М.В. Тувинские героические сказания. Свид. о гос. регистр. базы данных № 2017620090.

5. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP. СПб.: БВХ-Петербург, 2008. 384 с.

6. Быкова В.В., Монгуш Ч.М. Алгоритмы концептуального моделирования и классификации текстов в корпусе тувинского языка // Программные продукты и системы. 2017. Т. 30. № 3. С. 487–495.

7. Bykova V.V., Mongush Ch.M. On Algebraic Approach of R. Wille and B. Ganter in the Investigation of Texts // Journal of Siberian Federal University. Mathematics and Physics. 2017. no. 3(10). pp. 372–384.

8. Харламов А.А., Ермоленко Т.В., Дорохина Г.В. Сравнительный анализ организации систем синтаксических парсеров // Инженерный вестник Дона. 2013. № 4. URL: [ivdon.ru/ru/magazine/archive/n4y2013/2015](http://ivdon.ru/ru/magazine/archive/n4y2013/2015)

9. Кудинова Т.А. Языковой субстандарт в ткани художественных произведений современных писателей // Инженерный вестник Дона. 2012. № 3. URL: [ivdon.ru/ru/magazine/archive/n3y2012/1006](http://ivdon.ru/ru/magazine/archive/n3y2012/1006)

10. Соловьев С.Ю., Стельмашенко Д.Е. Подходы к исследованию формальных контекстов // Информационные процессы. 2011. Т. 11. № 2. С. 277–290.

11. Ganter B., Wille R. Formal Concept Analyses: mathematical foundations. Springer Science and Business Media, 2012. 284 p.

12. Монгуш Ч.М., Ондар М.В. База данных и средства создания контекстов для представления и анализа тувинского героического эпоса // Программные продукты, системы и алгоритмы. 2017. № 3. С. 1–6.

### References

1. Bavuu-Syuryun M.V. Novyye issledovaniya Tuvy. 2016. № 4. URL: [nit.tuva.asia/nit/article/view/610](http://nit.tuva.asia/nit/article/view/610)





2. Orus-ool S.M. Tuvinskiye geroicheskiye skazaniya (tekstologiya, poetika, stil') [Tuvan heroic legends (textual, poetic, style)]. M.: Maks press, 2001. 422 p.
3. Ondar M. V. Novyye issledovaniya Tuvy. 2016. № 4 URL: [nit.tuva.asia/nit/article/view/616](http://nit.tuva.asia/nit/article/view/616).
4. Bavuu-Syuryun M.V., Dalaa S.M., Mongush CH.M., Ondar M.V. Tuvinskiye geroicheskiye skazaniya [Tuvan heroic legends]. Svid. o gos. registr. bazy dannykh № 2017620090.
5. Barsegyan A.A., Kupriyanov M.S., Stepanenko V.V., KHolod I.I. Tekhnologii analiza dannykh: Data Mining, Visual Mining, Text Mining, OLAP [Data Analysis Technologies: Data Mining, Visual Mining, Text Mining, OLAP]. SPb.: BVKH-Peterburg, 2008. 384 p.
6. Bykova V.V., Mongush CH.M. Programmnyye produkty i sistemy. 2017. V.30. №3. pp. 487-495.
8. KHarlamov A.A., Ermolenko T.V., Dorokhina G.V. Inzhenernyj vestnik Dona (Rus). 2013. № 4. URL: [ivdon.ru/ru/magazine/archive/n4y2013/2015](http://ivdon.ru/ru/magazine/archive/n4y2013/2015)
9. Kudinova T.A. Inzhenernyj vestnik Dona (Rus). 2012. № 3. URL: [ivdon.ru/ru/magazine/archive/n3y2012/1006](http://ivdon.ru/ru/magazine/archive/n3y2012/1006)
10. Solov'yev S.YU., Stel'mashenko D.E. Informatsionnyye protsessy. 2011. V. 11. № 2. pp. 277–290.
12. Mongush CH.M., Ondar M.V. Programmnyye produkty, sistemy i algoritmy. 2017. № 3. pp. 1–6.