УДК 004.81+61

ПРОБЛЕМАТИКА ПОСТРОЕНИЯ ОНТОЛОГИЙ С ИЕРАРХИЧЕСКОЙ СТРУКТУРОЙ

Газизова М. Д. [1], Гольдштейн С. Л. [1], Донцов О. Г. [2]

[1] ФГАОУ ВПО «УрФУ», г. Екатеринбург, Россия
[2] ГАУЗ СО МКМЦ «Бонум», г. Екатеринбург, Россия

В статье приведен обзор методов автоматизированного построения предметных онтологий иерархического типа. Осуществлены анализ и критика аналогов по методологическим схемам и метрикам оценки качества онтологии. Обозначена применимость аналогов.

Ключевые слова: онтологии, иерархическая структура, построение и анализ онтологий.

The problems of construction of ontologies with a hierarchical structure

Gazizova M. D. [1], Goldstein S. L. [1], Dontsov O. G. [2]

[1] Ural Federal University, Ekaterinburg
[2] State Autonomous Health Institution Sverdlovsk Region Multidisciplinary Clinical Medical Center "Bonum", Ekaterinburg

The article presents a review of methods for automated generation of subject ontology the hierarchical type. The analysis and criticism unparalleled in the methodological schemes and metrics for assessing the quality of ontology. Indicated the applicability of analogues.

Keywords: ontologies, the hierarchical structure, the construction and analysis of ontologies.

Введение

Онтологии широко используют в представлении знаний, инженерии знаний, информационном поиске, семантической интеграции информационных ресурсов и др. Сформированные на основе онтологий интеллектуальные системы продуктивны в различных сферах деятельности, в том числе, в медицине [1-2]. Существует 3 способа построения онтологии: ручной, автоматизированный и автоматический. Способам создания онтологий посвящен ряд публикаций, например [3-13]. Разработаны программные продукты, наиболее известные: Protégé [12], Ontolingua[13], OntoEdit [14], Concept-Ont-M2 [4], CONCEPT MARKER [3] и др. При этом проблема автоматизации построения онтологий остается актуальной.

Заказчик текущей работы – учреждение системы здравоохранения ГАУЗ СО – многопрофильный клинический медицинский центр «Бонум».

В статье поставлена и решена задача анализа источников информации по теме, выбора аналогов автоматизированного построения онтологий, их сравнения с известными методологическими схемами, анализа и критики аналогов, уточнения области применения.

Литературно-аналитический обзор

В ходе данной работы были просмотрены 8 монографий и учебников, 25 научных статей и множество ссылок интернет. Выявлены методологические схемы построения онтологий [3 - 4]. Для дальнейшего анализа отобраны 4 аналога методов автоматизированного построения онтологий [5 -7, 13], удовлетворяющие условию: научная статья должна быть не старше 2007 г.

На основе литературно-аналитического обзора принято решение, прежде всего, представить исходные вербальные описания всех аналогов в виде алгоритмов на языке блок-схем.

Алгоритмизация аналогов на языке блок-схем

В разделе представлены восстановленные из авторского (иходного) вербального описания блок-схемы алгоритмов аналогов [5-7, 11] (рис.1 - 4). Видно, что строгость в отраженнии порядка действий и наглядность повысились по сравнению с вербальными описаниями.

Относительно термина «концепт» в блоке 4 рис. 1 поясним: концепт — единица речевого высказывания, логический и смысловой компонент его семантической структуры; характеризует акт понимания и его результат, полученный в коммуникации, тем самым предполагая направленность на другого; акт схватывания смыслов [14].

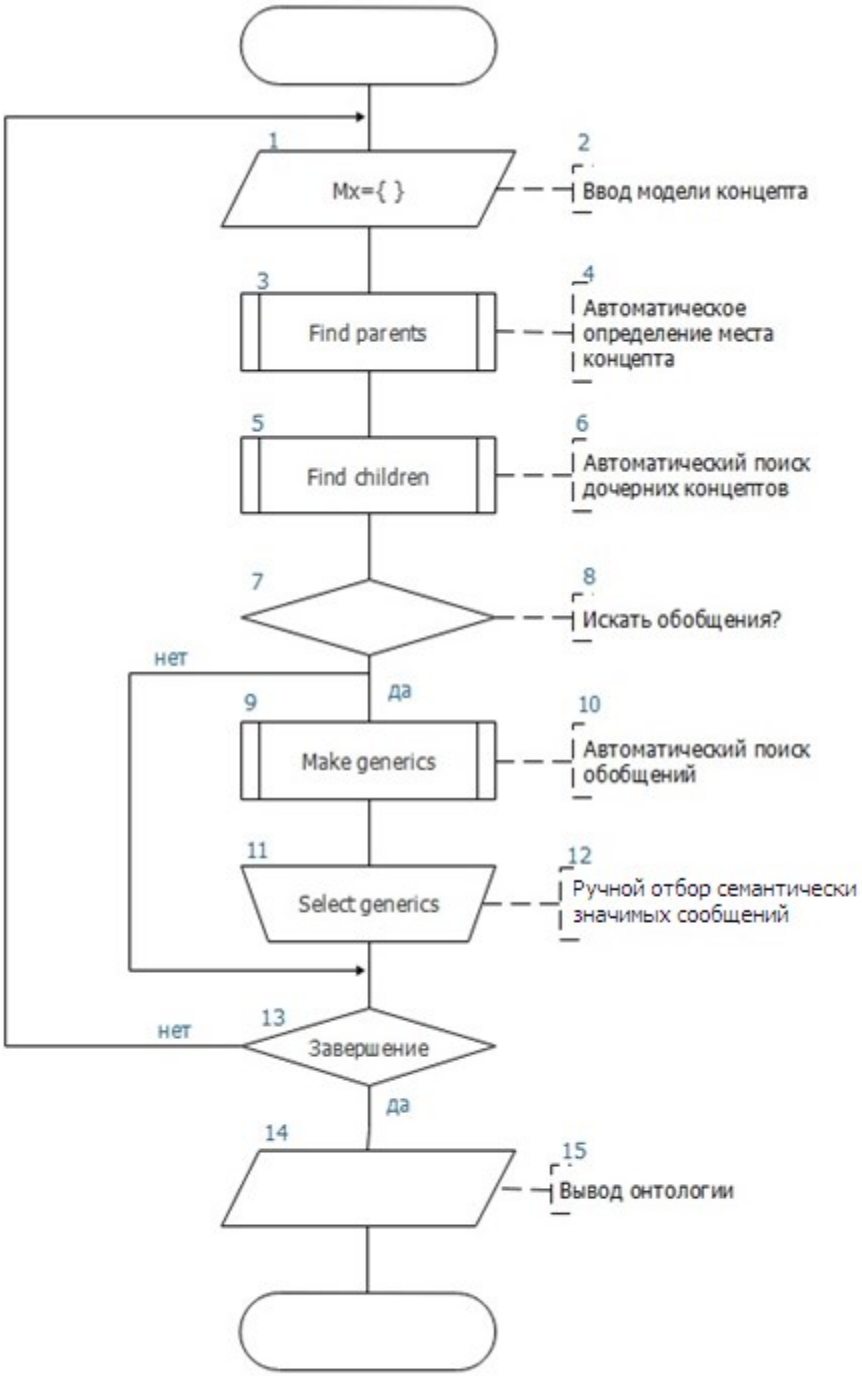


Рис. 1. Блок-схема метода автоматизированного построения онтологии предметной области [5]

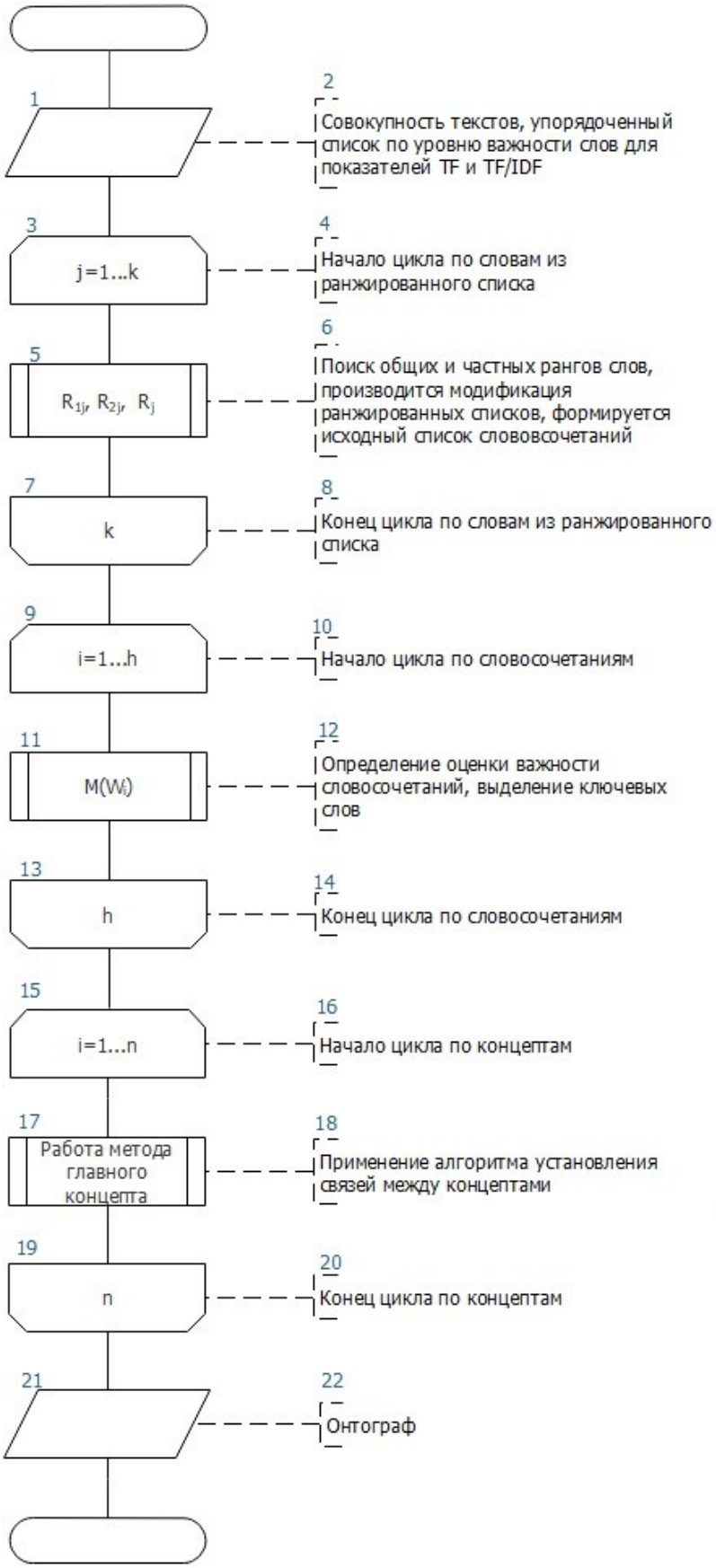


Рис. 2. Блок-схема метода автоматизированного построения онтологических моделей с древовидной структурой концептов [6]

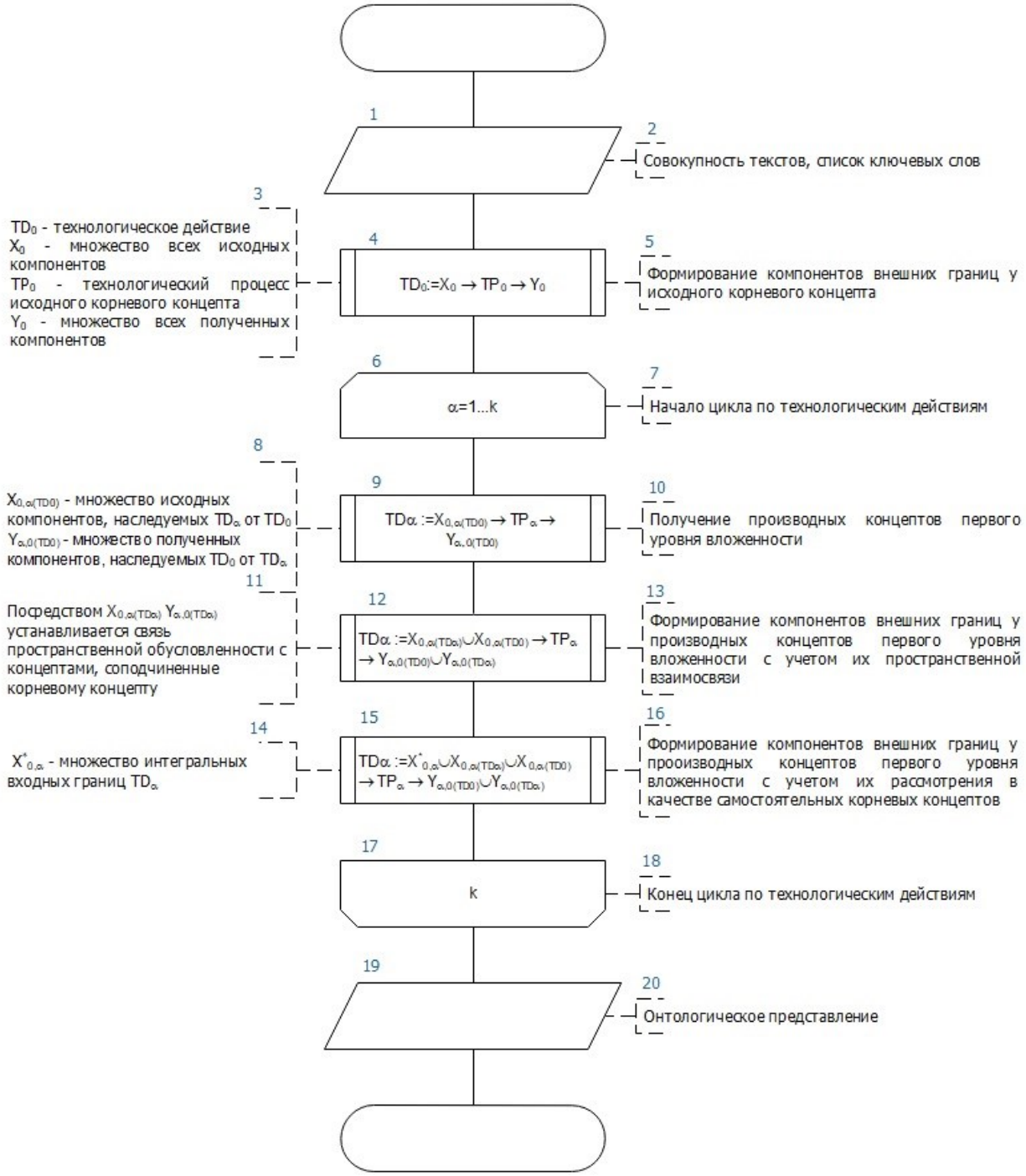


Рис. 3. Блок-схема метода построения онтологии технологических действий [7]

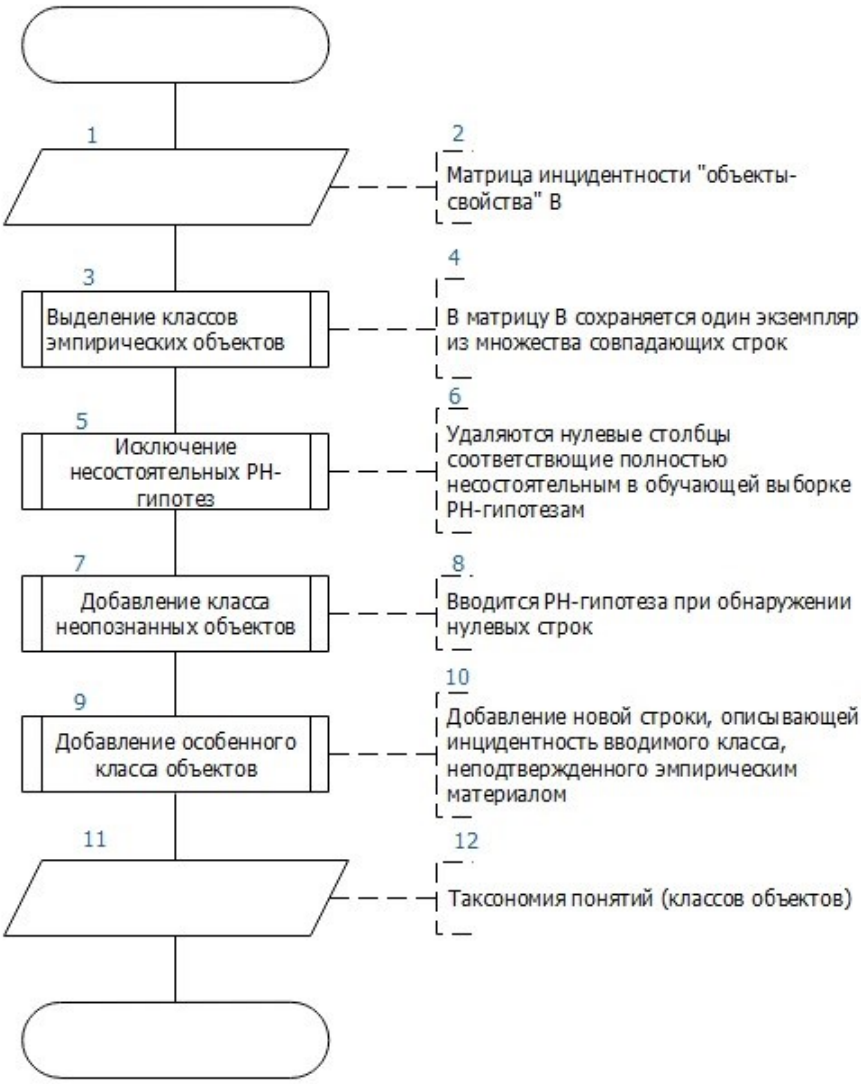


Рис. 4. Блок-схема метода построения онтологий предметных областей со структурными отношениями на основе анализа формальных понятий [13]

Методологические схемы как основа для анализа алгоритмов

В качестве основы для анализа алгоритмов взяты методологические схемы [3]. Основная приведена на рис. 5. Видно, что для построения онтологий предметной области предлагается ориентировка на терминосистемы (блок 1 из 4-х процедур: извлечение знаний о заголовочных терминах, извлечение дефиниций заголовочного термина, извлечение знаний о семантических отношениях и построение фрагментов фреймов-знаков, построение единой семантической сети фреймов-знаков) или на номенклатуру (блок 2 из 4-х процедур: извлечение дефиниций терминов, извлечение знаний о семантических отношениях, построение фрагментов фреймов-знаков, соединение номенклатуры и терминосистемы). При этом заложена предварительная обработка естественно-языкового текста традиционными методами по схеме на рис. 6. Построение онтологии, как терминосистемы, даёт ядро онтологии предметной области, а конечный

вариант онтологии должен быть создан путем соединения нескольких ядер, построенных на основе разных терминологических словарей. При построении онтологии как номенклатуры, онтология представляется иерархией, в корне которой - терминосистема, а в узлах – номенклатура. Сложность автоматизированного построения онтологии заключается в том, что велика роль субъективного начала в формировании концепта.

Алгоритм [5] не имеет собственного средства предварительной обработки естественно-языкового текста. Модели концептов предметной области вводятся вручную экспертом или из внешних источников в результате работы программного обеспечения. В алгоритме [6] предварительная обработка текстов основывается на статистическом анализе. При этом отсутствует четкое разделение между понятиями «слово» и «концепт». В алгоритмах [7, 13] способы предварительной обработки не конкретны.

Оценка качества построения онтологий

В табл. приведено сравнение выбранных аналогов по известным метрикам оценки онтологий [15]. Можно отметить, что в алгоритмах всех аналогов отсутствует такая проверка.

Таблица
Сравнение аналогов методов автоматизированного построения онтологии

Ссылка	Оценки по критериям:							Общая оценка
	общего анализа графа	когнитивной эргономичности	метрики циклов	Ингве-Миллера	разнообразия количества связей	глубины	запутанности графа	
[5]	0,55	1	1	1	1	0	0,9	5,45
[6]	0,64	0,42	1	1	0,33	0,37	0,89	4,65
[7]	1	0,84	1	1	1	0,43	1	6,27
[13]	0,82	0,97	1	1	Нет данных	0,79	1	5,58



Рис. 5. Обобщенная схема построения онтологии предметной области [3]

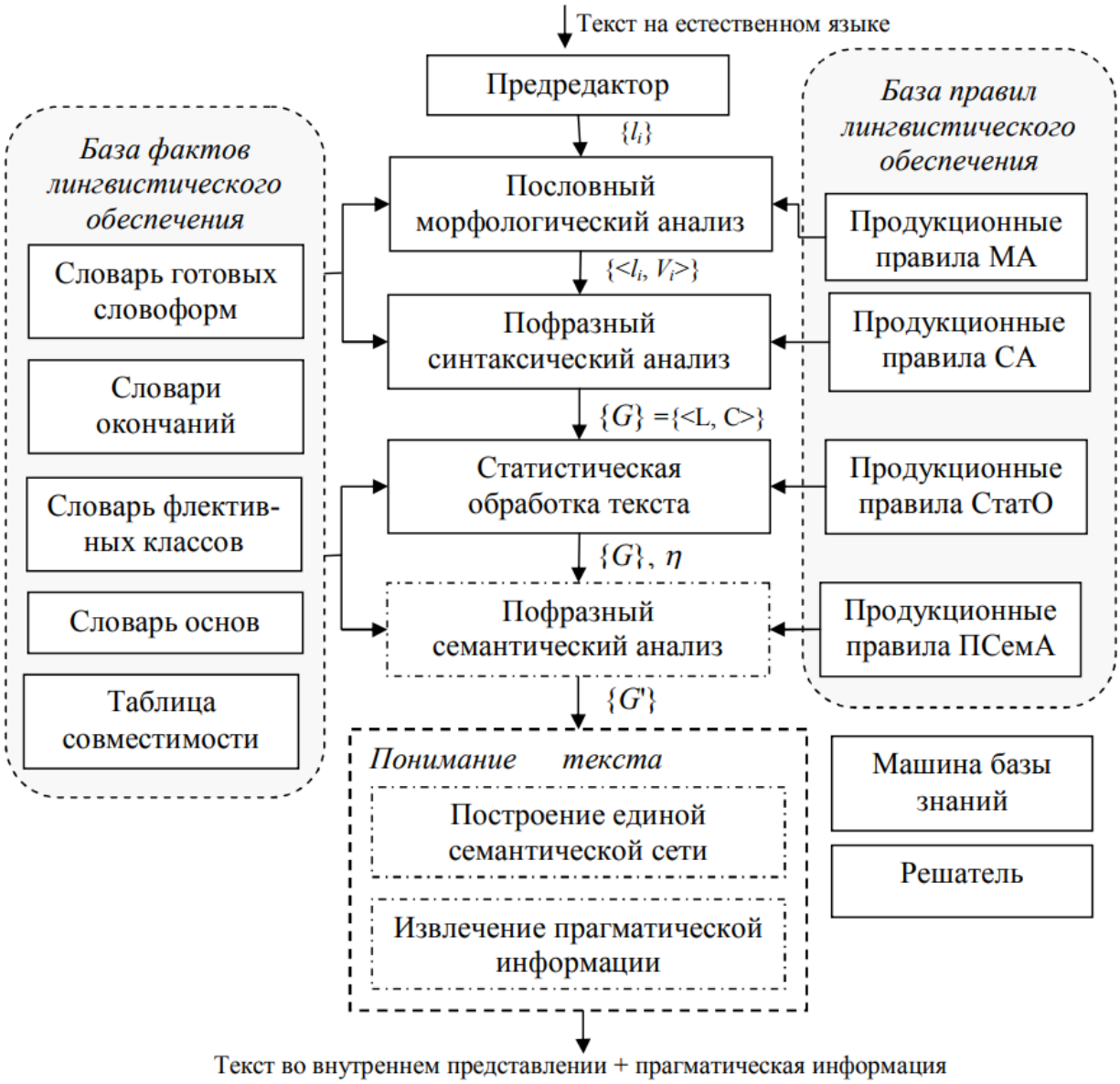


Рис. 6. Обобщенная схема анализа монологического текста [3]

Применимость алгоритмов аналогов

Методы, выбранные в качестве аналогов, можно применять к любым текстам. Задача нашей дальнейшей работы – применить к деятельности медицинской организации в части собственно медицины (диагностика, лечение, реабилитация, профилактика), организации здравоохранения и вспомогательных технологий (эпидемиология, IT-технологии), в частности, разобраться с качеством разработки программных средств для медицинских информационных систем, например, по стандартам ГОСТ и ISO.

Результаты и выводы

1. Поставлена задача анализа источников информации о методах автоматизированного построения онтологий, выбора аналогов, их сравнения с известными методологическими схемами, анализа и критики аналогов, уточнения области применения,

2. Из вербального описания методов восстановлены алгоритмы аналогов на языке блок-схем,

3. Проведен анализ 4-х аналогов на основе известной методологической схемы, проведена критика аналогов,

4. Осуществлена оценка качества онтологий на основе известных критериев,

5. Определена область применения

В дальнейшем целесообразно выбрать компилятивный прототип и поставить эксперимент с построением иерархической онтологии по прототипу и аналогам.

Список литературы

1. Клещев А.С. Математические модели онтологий предметных областей. Часть 1. Существующие подходы к определению понятия «онтология» / А.С. Клещев, И.Л Артемьева // Научно-техническая информация, серия 2 «Информационные процессы и системы». – 2001. – № 2. – С. 20-27..

2. Humphreys B.L. The Unified Medical Language System: An Informatics Research Colloboration / B.L. Humphreys, D.A.B. Lindberg, H.M. Schoolman, G.O.Barnett // J. Am. Medical Informatics Assoc. – 1998. – Vol. 5. – No. 1. – P. 1-11.

3. Нахайнова Л. В. Технология создания методов автоматического построения онтологий с применением генетического и автоматного программирования / Л. В. Нахайнова. – Улан-Удэ : БНЦ СО РАН, 2008. – 236 с.

4. Ермаков А. Е. Автоматизация онтологического инжиниринга в системах извлечения знаний из текста / А. Е. Ермаков. // Труды Международной конференции Диалог. – 2008.

5. Антонов И. В. Метод автоматизированного построения онтологии предметной области : дис. … канд. техн. наук : 05.13.01 / И. В. Антонов; Балт. гос. техн. ун-т. – 2011. – 156 с.

6. Чалая Л. А. Метод автоматического построения онтологических моделей с древовидной структурой концептов / Л. А. Чалая, А. В. Чижевский. // Автоматизированные системы управления и приборы автоматики. – 2015. – 173. – С. 32-42

7. Андреев Д. А. Метод построения онтологии технологических действий / Д. А. Андреев, М. В. Воронов. // Вестник СГТУ. – 2012. – №3(67). – С. 160-168.

8. Кравченко Ю. А. Метод создания онтологии предметной области на основе глоссария / Ю. А. Кравченко, А. А. Новиков, В. В. Марков. // Известия Южного федерального университета. Технические науки. – 2015. – 6(167) – С. 158-168.

9. Оробинская Е.А. Метод автоматического построения онтологии предметной области на основе анализа лингвистических характеристик текстового корпуса / Е.А. Оробинская // Интернет и современное общество (IMS-2012) : тр. XV Всерос. объединенной конф., СПб., 2012. – С. 209-212.

10. Грегер, С. Э. Построение онтологии архитектуры информационной системы / С. Э. Грегер, С. В. Поршнев. // Фундаментальные исследования. – 2013. – 10(11). – С. 2405-2409. – URL: https://fundamental-research.ru/pdf/2013/10-11/32804.pdf

11. Gavrilova T. Gestalt principles of creating learning business ontologies for knowledge codification / T. Gavrilova, I. Leshcheva, E. Strakhovich. // Knowledge Management Research & Practice. – 2015. – 13(4). – С. 418-428.

12. Gavrilova T. A. Ontology design and individual cognitive peculiarities: A pilot study / T. A. Gavrilova, I. A. Leshcheva, E. Strakhovich. // Expert systems with Applications. – 2015. – 42(8). – С. 3883-3892.

13. Смирнов С.В. Построение онтологий предметных областей со структурными отношениями на основе анализа формальных понятий / С.В. Смирнов // Знания – Онтологии - Теории: Труды Всероссийской конференции с международным участием ЗОНТ-2011. - Новосибирск : Институт математики СО РАН, 2011. — Т. 2. — С. 103-112.

14. Касавин И. Т. Энциклопедия эпистемологии и философии науки / И. Т. Касавин. – М. : «Канон+», РООИ «Реабилитация», 2009. – 1248 с.

15. Гаврилова, Т. А. Субъективные метрики оценки онтологий / Т. А. Гаврилова. и др. // Знания – Онтологии – Теории : Материалы Всероссийской конференции с международным участием ЗОНТ-09 . – 2009. – С. 178-187.

16. Musen M. Domain Ontologies in Software Engineering: Use of Prot?g? with the EON Architecture / M. Musen. // Methods of Inform. in Medicine. – 1998. – 37(14). – С. 540-555.

17. Farquhar A. The Ontolingua server: A tool for collaborative ontology construction / A. Farquhar, R. Fikes, J. Rice. // International Journal of Human-Computer Studies. – 1997. – 46(6). – С. 707-728.

18. Sure Y. OntoEdit: Collaborative ontology development for the Semantic Web / Y. Sure. и др. // In Proc. of the Inter. Semantic Web Conference (ISWC 2002). – 2002. – URL: https://link.springer.com/chapter/10.1007/3-540-48005-6_18

19. Соловьев В. Д. Онтологии и тезаурусы / В. Д. Соловьев, Б. В. Добров, В. В. Иванов, Н. В. Лукашевич – Казань: Казанский государственный университет, 2006. – 157 с.

20. Гольдштейн С. Л. Настройка корпоративных информационных систем на задачи предприятия / Гольдштейн С. Л., Кашперский И. В. ; ГОУ ВПО "Уральский гос. технический ун-т - УПИ", НП "Уральский межакадемический союз". - Екатеринбург : Форт Диалог-Исеть, 2006. – 132 с.

Газизова Мария Дамировна – бакалавр кафедры технической физики, физико-технологического института ФГАОУ ВО УрФУ им. Первого президента России Б.Н. Ельцина, 620002, Екатеринбург, ул. Мира, 19, тел. +7 (343) 375-41-51, gazizova-mashka.gazizova@yandex.ru