

**Approaches to develop domain vocabulary based on work breakdown structure  
Danilova V.<sup>1</sup>, Kshnyakov D.<sup>2</sup>, Ananieva A.<sup>3</sup>, Suslina A.<sup>4</sup>**

**О подходах к разработке словаря предметной области на основе иерархической  
структуры работ**

**Данилова В. В.<sup>1</sup>, Кшняков Д. О.<sup>2</sup>, Ананьева А. Г.<sup>3</sup>, Суслина А. С.<sup>4</sup>**

<sup>1</sup>Данилова Валерия Викторовна / Danilova Valeriya Viktorovna – студент;

<sup>2</sup>Кшняков Дмитрий Олегович / Kshnyakov Dmitriy Olegovich – аспирант;

<sup>3</sup>Ананьева Анастасия Геннадьевна / Ananieva Anastasia Gennadievna – ассистент;

<sup>4</sup>Суслина Алена Сергеевна / Suslina Alyona Sergeevna – студент,

кафедра анализа конкурентных систем,

Национальный исследовательский ядерный университет,

Московский инженерно-физический институт, г. Москва

**Аннотация:** рост количества публикуемой тематической информации ставит перед исследователями необходимость разработки новых подходов к построению алгоритмов фильтрации больших потоков входящей информации. В статье рассматривается подход формирования словаря ключевых слов по направлению «Специальные разработки» на основе иерархической структуры работ (ИСР). Разработка данного словаря входит в комплекс работ по созданию Мультиагентной информационно-аналитической системы по естественно-научным и технологическим направлениям.

**Abstract:** problem of filtering large streams of incoming information is discussed in the article. Constantly increasing number of information items (i.e. published or digital subjects) set the necessary goal to develop new approaches for filtering and structuring such data. The article discusses the approach of forming a dictionary of keywords in the direction of "Special study" based on US "Work breakdown structures for defense materiel items" (WBS). The development of this dictionary is part of a work on realization of Multi-agent information and analytical system for scientific and technological directions.

**Ключевые слова:** мультиагентная система; иерархическая структура работ; ИСР.

**Keywords:** multiagent system; work breakdown structure; WBS.

В настоящее время в различных отраслях науки существует необходимость работы с большими объемами информации. Обработка таких массивов невозможна без разработки специализированных иерархических или сетевых моделей соответствующих предметных областей.

Одной из таких моделей описания предметной области является иерархическая структура работ (ИСР) или Work breakdown structure (WBS). Понятие WBS было разработано в рамках Программы оценки и контроля проекта (Program Evaluation and Review Technique (PERT)) министерством обороны США [1].

В данной работе исследуется возможность построения словаря ключевых слов для описания предметной области «Специальные разработки» на основе открытых материалов.

Основным документом для построения словаря ключевых слов стал документ Министерства обороны США "Work breakdown structures for defense materiel items" (MIL-STD-881C). Он представляет собой инструкцию для эффективной подготовки, понимания и представления иерархической структуры работ (ИСР) в области разработки оборонной продукции. Она состоит из последовательных общих принципов для организации и структурирования оборонительных средств, также включает в себя унифицированные определения и логическую последовательность в построении всех уровней декомпозиции. Применение единообразной структуры способствует улучшению информационного взаимодействия в процессе исследования, разработки, испытаний и оценки продукции.

Целью документа является унификация всех этапов жизненного цикла тематического направления «Специальные разработки» таких как:

1. Производительность;
2. Стоимость;
3. Сроки;
4. Риски;
5. Бюджет.

Первой важной целью создания ИСР являлась разработка такой структуры, которая определяла бы логическую связь между всеми элементами до определённого уровня (как правило, до 3 или 4 уровня) по выполняемой работе, которая не будет ограничивать возможности регулирования и управления программами и ресурсами.

Второй важной целью являлось обеспечение систематического и стандартизованного метода для сбора данных по всем категориям, что позволило иметь актуальные данные для мониторинга аналогичных материально-технических средств.

В том числе ИСР выступает в качестве координирующей среды (инфраструктуры) для обобщения данных, а также предоставления соответствующей информации о текущем и планируемом состоянии отдельных элементов. При строгом структурировании и использовании в сочетании с принципами системной инженерии, составлением смет, интегрированным планированием и управлением рисками, ИСР позволяет видеть полную картину о состоянии реализации программы, и на её основе вносить необходимые изменения для достижения желаемого результата.

В течение жизненного цикла программы или проекта ИСР выполняет следующие функции:

1. делит предметы снабжения оборонной продукции на составные части, выясняя связь между частями, взаимосвязь задач, которые должны быть выполнены и привести к конечному продукту;
2. способствует эффективному планированию и распределению управленческих и технических обязанностей;
3. отслеживает текущий статус расходов, распределения ресурсов, технических усилий, технических характеристик.

В каждой ИСР должно содержаться, по крайней мере, три верхних уровня, но также возможно увеличение уровней, например, до пяти. Рассмотрим взаимосвязь уровней на примере тематики «Авиационный комплекс» (**Ошибка! Источник ссылки не найден.**).

Уровень 1 включает себя всю систему и/или программу, программный элемент, проект или подпрограмму, например, «Aircraft System» – авиационный комплекс.

Элементы Уровня 2 являются основными в подчинении элементам Уровня 1. Например, «air vehicle» (летательный аппарат), является подразделением раздела авиационного и ракетного комплексов. Данный уровень также включает в себя комплекс данных и системного обслуживания (например, системная разработка, её испытание и оценка).

Элементы Уровня 3 подчиняются элементам Уровня 2, они включают в себя, например, аппаратные средства и программное обеспечение. Например, «процессор радиолокационных данных» является подчинённым элементов раздела «прицельная РЛС», или «доводочные испытания (испытания на стадии разработки) и их оценка» является подчинённым элементом раздела «испытание системы и её оценка».

Элементы Уровня 4 проходят тот же процесс, что и элементы Уровня 3 и представляют собой дальнейшее определение аппаратных средств, программного обеспечения.

На основе анализа приведённого документа был составлен словарь ключевых слов по направлению «Авиационный комплекс» (Aircraft System) для фильтрации входящей информации.

Составленный словарь включает в себя 30 терминов. Были проведены экспериментальные исследования по оценке качества информационной выдачи в мультиагентной информационно-аналитической системе по естественно-научным и технологическим направлениям (МИАС) [2, 3] по таким критериям, как точность соответствия материалов в информационной выдаче структурированному словарю по тематике «Компоненты авиационного комплекса». В качестве исходного массива информации был взят месячный объём информации, принесённый роботами в систему за февраль 2016 года.

Результаты эксперимента представлены в таблице

Таблица 1. Результаты эксперимента

№	Период мониторинга	Количество поступивших сообщений	Отфильтровано по словарю «Компоненты авиационного комплекса»	Контрольная выборка
1.	1 неделя	689	290	25
2.	2 неделя	707	315	25
3.	3 неделя	929	329	20
4.	4 неделя	934	342	16
<b>ИТОГО</b>		<b>3259</b>	<b>1276</b>	<b>86</b>

Как видно из таблицы составленный словарь позволил практически в 3 раза сократить объём входящей информации. Необходимо отметить, что первоначальный набор информационных материалов уже был получен с тематических ресурсов по приведенному выше направлению.

В результате автоматической фильтрации было отобрано всего 1276 документов, причем, в каждом из них присутствовало не менее 1 слова из словаря «Компоненты авиационного комплекса».

Также было доказано, что ни один документ из оставшейся выдачи (3259) не содержал информации о компонентах авиационного комплекса.

На примере случайно отобранных материалов (контрольная выборка) было показано, что в 86 документах содержится информация о компонентах авиационного комплекса. Пример информации из контрольной выборки в Таблице 2.

Таблица 2. Информация из контрольной выборки

Название документа	Дата публикации	Текстовый массив (фраза)	Термин из словаря компонентов авиационного комплекса	Оценка экспертом соответствия документа области компонентов авиационного комплекса
NASA wants to bring back X-planes to test new aviation technologies	21.02.2016	Beyond individual technologies, the X-planes will include novel designs, such as a truss-braced wing for subsonic flight	wing	Подходит
Bugatti Chiron revealed at Geneva 2016: the world has a new fastest production car	29.02.2016	There are two cooling water loops (high temp for engine and low temp for charge air cooling)	engine	Не подходит

Таким образом, можно сделать вывод о том, что словарь «Компоненты авиационного комплекса» является полным для описания данной области исследования. В дальнейшем планируется разработка словарей, позволяющих определять принадлежность информационных массивов определенным группам описания частей авиационного комплекса.

### Литература

1. Fleming, Quentin W., Joel M. Koppelman. "Earned Value Project Management" CROSSTALK: The Journal of Defense Software Engineering July 1998, p 20.
2. Артамонов А. А., Леонов Д. В., Оныкий Б. Н., Проничева Л. В. Мультиагентная информационно-аналитическая система по естественно-научным и технологическим направлениям // Системы высокой доступности, Т. 10, № 2, 2014.
3. Ananieva A. G., Artamonov A. A., Galin I. U., Tretyakov E. S., Kshnyakov D. O. Algoritmization of search operations in multiagent information-analytical systems // Journal of Theoretical and Applied Information Technology. Nov., 2015. Vol. 81. №. 1. pp. 11-17.
4. Salton G., McGill M. J. Introduction to modern information retrieval, 1986.