



УДК 004.8

© 2007 г. **Е.А. Шалфеева**, канд. техн. наук
(Институт автоматизации и процессов управления ДВО РАН, Владивосток)

МЕТОДЫ ОЦЕНИВАНИЯ СТРУКТУРНЫХ СВОЙСТВ ОНТОЛОГИЙ, РЕДАКТИРУЕМЫХ И СОХРАНЯЕМЫХ КОМПЬЮТЕРНЫМ БАНКОМ ЗНАНИЙ¹

Для пользователей компьютерного банка знаний важны свойства создаваемых и хранимых в нем онтологий и других теорий. В статье представлен подход к определению структурных свойств для объективного оценивания ресурсов банка. Для каждой теории банка можно построить разные внутренние графовые модели, отражающие ее различные структурные свойства. Значения этих свойств определяются на основе каталогизированных определений свойств, выраженных в терминах соответствующих графов. Такие свойства полезны при оценивании полноты, сложности, корректности редактируемых теорий.

Введение

Для исследования возможностей коллективного развития повторно используемых информационных ресурсов различных уровней общности разработан Многоцелевой банк знаний, который предоставляет, в частности, инженерам знаний проблемно-независимые средства с удаленным доступом для описания искусственных языков и онтологий, а экспертам – проблемно-ориентированные средства с удаленным доступом для создания и редактирования баз знаний и данных.

С помощью этих средств созданы версии некоторых проблемно-независимых и проблемно-ориентированных информационных ресурсов.

Для оценивания онтологий и других информационных ресурсов известны такие методы оценивания как FIGO, OntoMetric, EvaLexon, Natural

¹ Работа выполнена при финансовой поддержке РФФИ, проект 06-070-89071-а "Исследование возможностей коллективного управления в семантическом вебе информационными ресурсами различных уровней общности", и ДВО РАН в рамках Программы №15 ОЭМПУ РАН, проект 06-И-П15-055 "Синтез интеллектуальных систем управления базами знаний и базами данных при управлении сложными объектами в условиях неопределенности".

Language Application metrics, OntoClean [1 – 3], однако они малоприменимы для изучения свойств онтологий, хранимых в банке. Методом, обеспечивающим более широкий спектр измеряемых свойств, является единый подход к оцениванию структурных свойств онтологий [4].

Суть подхода состоит в том, что для всякой онтологии можно построить разные внутренние модели, отражающие различные внутренние (структурные) свойства онтологии, а для представления таких моделей – использовать размеченные графы. Предлагаемые в рамках единого подхода графовые модели онтологий разбиты на следующие группы: графы синтаксических связей, графы стандартных связей, графы концептуальных связей и графы проблемно-ориентированных связей. Сформированный каталог свойств онтологий [4, 5] содержит наборы однозначных определений свойств каждого класса (т.е. по каждой графовой модели из каждой группы). Свойства структуры онтологии определяются в терминах соответствующих графов через подсчет числа вершин или дуг, числа одноименных дуг, разветвлений, числа и доли дуг с определенными метками, плотности связей среди нескольких вершин, наличие вершин или путей, характеризующихся заданными признаками, длин путей и др.

Цель работы – показать применение единого подхода к оцениванию структурных свойств для анализа свойств онтологий и других теорий, хранимых в банке.

Постановка задачи

Для разработки метода оценивания информационных ресурсов различных уровней общности, хранимых Многоцелевым банком знаний, предлагается использовать единый подход к оцениванию структурных свойств, при этом требуется из множества предлагаемых [4] графовых моделей выбрать наиболее содержательные и уточнить правила построения на их основе внутренних представлений для теорий банка.

Уточнение правил построения графовых моделей позволит учесть особенности языка представления («информация общего вида на языке ИРУО представляет собой семантическую сеть понятий – набор понятий и направленных отношений между ними, снабженных специальной разметкой, ... понятия сети представляют термины этого языка, а направленные дуги описывают, как одни термины языка определены через другие» [6]). Кроме того, следует уточнить набор определений структурных свойств, который актуален для оценивания такого рода ресурсов.

Анализ особенностей языка представления теорий в банке приводит к рассмотрению в первую очередь *графов стандартных связей*. Целесообразно также использовать для оценивания *графы синтаксических связей* (они обеспечивают объективный анализ связей между теориями и между терминами в теориях – использованность одними компонентами других, наличие циклических связей), а от двух других групп графовых моделей эф-

фекта не ожидается.

Анализ партономических связей

Для анализа теорий, представленных на языке ИРУО, применим *граф стандартной партономии сущностей*. Этот граф из группы *графов стандартных связей* более других пригоден для измерения структуры теорий, поскольку все виды разметки семантической сети укладываются в спектр отношений «часть–целое». Внутренние (графовые) представления объективно строятся для хранимых в банке теорий, описанных с использованием конструкций целое-часть (является множеством, является последовательностью).

Граф стандартной партономии [модуля] онтологии – графовая модель $\langle B, D \rangle$,

где вершины $B = \{v_i\}$, v_i – составные и простые сущности (понятия) теории,

а дуги $D = \{d_j\}$, d_j – связи сущностей следующих видов: «включает часть» (является конечной последовательностью разнотипных элементов), «включает элемент» (является множеством), «один из» (имеет множество альтернатив), «есть» (считается эквивалентным). Дополнительно партономическая связь характеризуется кардинальностью (1 или *) и модальностью (0 или 1).

О методе построения *графа стандартной партономии*: каждому термину хранимой теории (узлу семантической сети) соответствует вершина; каждому представлению-разбиению термина в виде *последовательности терминов* соответствует совокупность дуг «включает часть», представлению термина как *множества терминов (конечного или бесконечного множества терминов)* соответствует совокупность дуг «включает элемент», представлению термина в виде *альтернативных терминов* соответствует совокупность дуг «один из», описанию термина как *объекта ИРУО* соответствует дуга «есть».

При построении графа для теории, определенной в терминах конструктивного базиса языка ИРУО (базовой теории, предназначенной для построения других теорий), исходят из того, что имена терминов уникальны: один и тот же термин может встречаться как составная часть/элемент/альтернатива различных других терминов. При построении графа для теории, использующей базовую теорию в качестве входной, более целесообразен другой прием. Имена терминов рассматриваемой/оцениваемой теории уникальны, а используемые имена терминов из входной теории будут дублироваться столько раз, сколько применяются.

Например, для теории «язык прикладной логики» из раздела «концептуальные знания» банка медицинских знаний² (назовем ее теорией 1)

² <http://mpkbank.dvo.ru/index.php?page=banks>

«терм-множество» представлен как одно из альтернативных понятий: «термин», «множество всех подмножеств», «интервалы» и др., а «множество всех подмножеств» определяется через терм-множество». В графе будет одна вершина, соответствующая понятию «терм-множество». Для теории «простая онтология медицины» (назовем ее теорией 2) понятия представляются с использованием понятия «терм-множество» (из входной теории) в роли одной из альтернатив. В графе будет столько вершин, соответствующих понятию «терм-множество», сколько раз рассматривалась какая-либо его альтернатива.

Модель позволяет определить множество структурных свойств, определения которых представлены в [4, 5]. Рассмотрим те из них, которые существенны для анализа хранимых в банке ресурсов.

С трудоемкостью проверки определений понятий (в частности после внесения изменений в те определения понятий, от которых зависят многие другие), связаны следующие структурные свойства.

Размер партономии.

Число вершин в графе стандартной партономии сущностей модуля онтологии.

Область значений: целое неотрицательное число.

Число корней партономии

Число корневых вершин в графе стандартной партономии.

Область значений: целые неотрицательные.

Глубина партономии.

Максимальная длина цепочки дуг (без циклов) от корневой вершины к листовой в графе стандартной партономии сущностей.

Область значений: целое неотрицательное число.

«Разветвление» узлов графа

Максимальное число дуг, тип которых не равен «один из», выходящих из вершины графа стандартной партономии сущностей.

Область значений: целое неотрицательное число.

«Альтернативность» узлов графа.

Максимальное число дуг «один из», выходящих из вершины графа стандартной партономии сущностей.

Область значений: целое неотрицательное число.

С корректностью представления совокупности понятий связаны следующие структурные свойства.

«Перекрываемость частей» – нарушение древовидности при де-

лении целого на части.

Наличие в *графе стандартной партономии* вершин с полустепенью захода больше единицы. (Если граф строится без требования к уникальности имен вершин, то свойство вычисляется как максимальное число вхождений некоторого имени в качестве вершины графа.)

Область значений: одно из значений {да, нет}.

Наличие циклической связи «часть-целое» (сущность состоит из самой себя или сущность 1 состоит из сущности 2 и наоборот)

Наличие цепочек дуг в *графе стандартной партономии*, начинающихся и заканчивающихся одной и той же вершиной.

Область значений: одно из значений {да, нет}.

«Двойная роль» компонента сущности.

Наличие двух путей от вершины *B1* к вершине *B2* в *графе стандартной партономии* – таких, что типы начальных либо конечных дуг этих путей различны.

Область значений: одно из значений {да, нет}.

Нарушение баланса партономии.

Модуль разности длин самого длинного и самого короткого пути в графе в *графе стандартной партономии сущностей*.

Область значений: целое неотрицательное число.

Повторяемые альтернативные подгруппы – Число \ наличие групп терминов, которые являются одновременно альтернативами двух или более сущностей.

Совокупность имен вершин в *графе стандартной партономии сущностей*, в каждую из которых входят две дуги «один из» из одинаковых двух вершин-сущностей.

Область значений: множество множеств имен.

Максимальный линейный участок.

Максимальная длина (в дугах) цепочки, связывающей вершины (кроме последней в цепочке) с полустепенью исхода, равной единице, дугами с кардинальностью, равной единице.

Область значений: целое неотрицательное число.

Рассмотрим примеры измерения теорий банка в соответствии с приведенными выше определениями структурных свойств.

Пример 1. В табл. 1 приведены значения свойств структуры партономических связей, существенные для фрагмента теории 1, представленного на рис. 1.

Таблица 1

Название свойства	Значение свойства
Размер партономии модуля	30
Глубина партономии	9 дуг
Число корней партономии	1
«Альтернативность» узлов графа»	11
«Перекрываемость частей»	Да
Наличие циклической связи «часть-целое»	да
Повторяемые альтернативные подгруппы	{{Термин, Переменная, Условный терм} (альтернативы для «терм», Терм-МН-ВО», «Целочисл терм»), {Сумма чисел, Разность чисел, Произв чисел, Частное от деления, Возв в степ, Константа} (- альтернативы для «Ариф термы», «Целочисл терм»)}
Максимальный линейный участок	1
Баланс партономии	7
«Двойная роль» компонента сущности	нет
«Разветвление» узлов графа	3

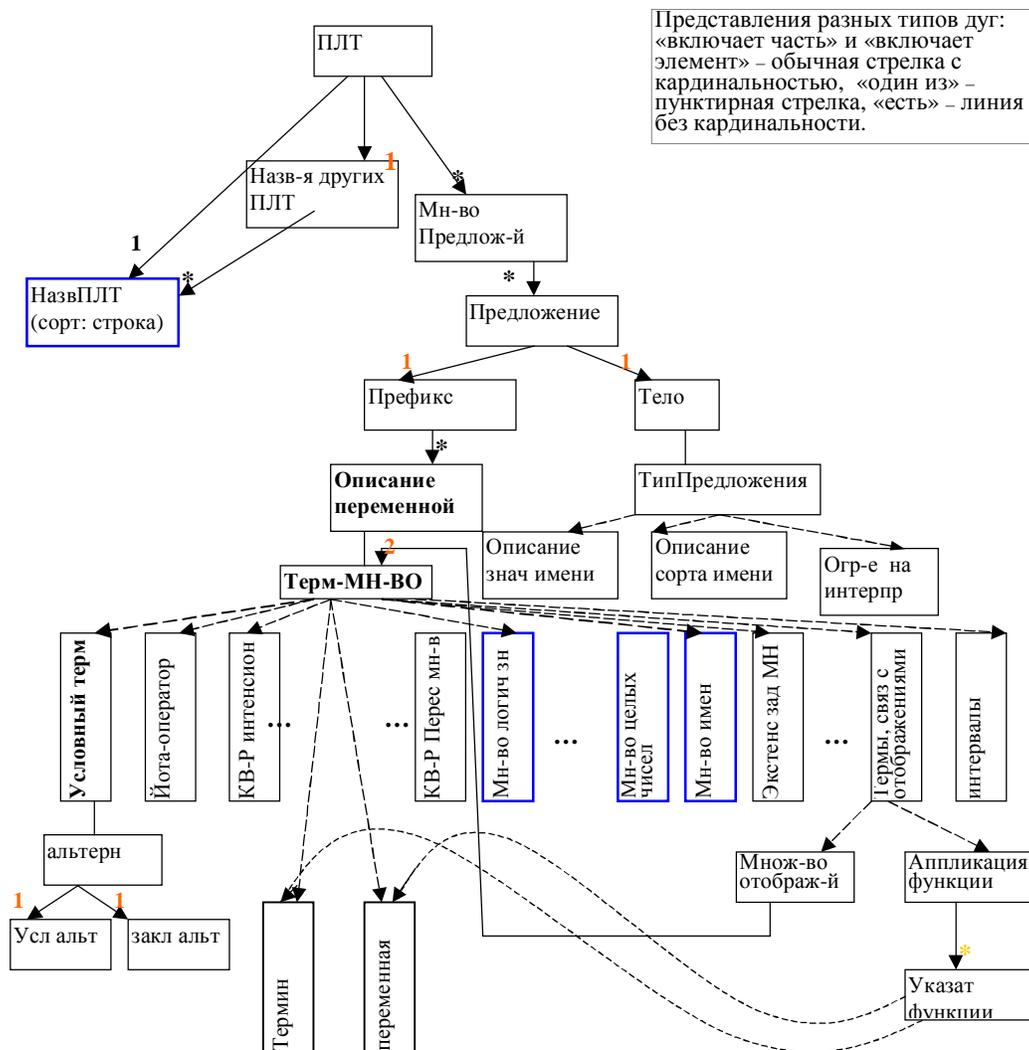


Рис. 1. Фрагмент графа партономических связей сущностей теории 1.

На основании этих значений можно сделать предварительные выводы о потенциальной сложности самой теории и теорий, которые будут использовать ее как входную.

Линейные участки теоретически могут быть заменены на вершины, что может уменьшить размер теории. А наличие *повторяемых альтернативных подгрупп* может служить рекомендацией к выделению их в отдельные понятия, что, однако, может привести к увеличению размера теории. Цикличность для базовой теорий означает, что теории, которые будут использовать ее как входную, могут содержать цепочки очень большой длины.

Пример 2. В табл. 2 приведены значения свойств структуры партономических связей, существенные для фрагмента теории 2 (рис. 2).

Таблица 2

Название свойства	Значение свойства
<i>Размер партономии модуля</i>	37
<i>Глубина партономии</i>	15 дуг
<i>«Разветвление» узлов графа</i>	4
<i>Число корней партономии</i>	1
<i>«Перекрываемость частей»</i>	Да
<i>Наличие циклической связи «часть-целое»</i>	нет
<i>«Двойная роль» компонента сущности</i>	нет
<i>Нарушение баланса партономии</i>	5
<i>Максимальный линейный участок</i>	1

Для теорий, сформированных в терминах некоторой другой (входной) теории, свойства «Альтернативность узлов графа» и «Повторяемые альтернативные подгруппы» уже не несут полезной информации, так как всегда равны соответственно единице и пустому множеству.

На основании значений, приведенных в табл. 2, можно сделать выводы о влиянии сложности входной теории на теорию, которая использовала ее как входную.

Анализ «теоретико-множественных» связей

Другим применимым для представления структуры теорий графом является *граф теоретико-множественных связей*. В данном контексте он является подграфом предыдущей графовой модели и охватывает только сущности, связанные отношением «является множеством».

Граф «теоретико-множественных» связей [модуля] онтологии – графовая модель $\langle V, D \rangle$,

где вершины $V = \{v-i\}$, $v-i$ – составные и простые сущности (термины) теории,

а дуги $D = \{d-j\}$, $d-j$ – связи сущностей следующих видов: *включает элемент* (т.е. термин *является множеством из других терминов*) и «альтернативное множество» (т.е. термин *является одним из* указанных терминов).

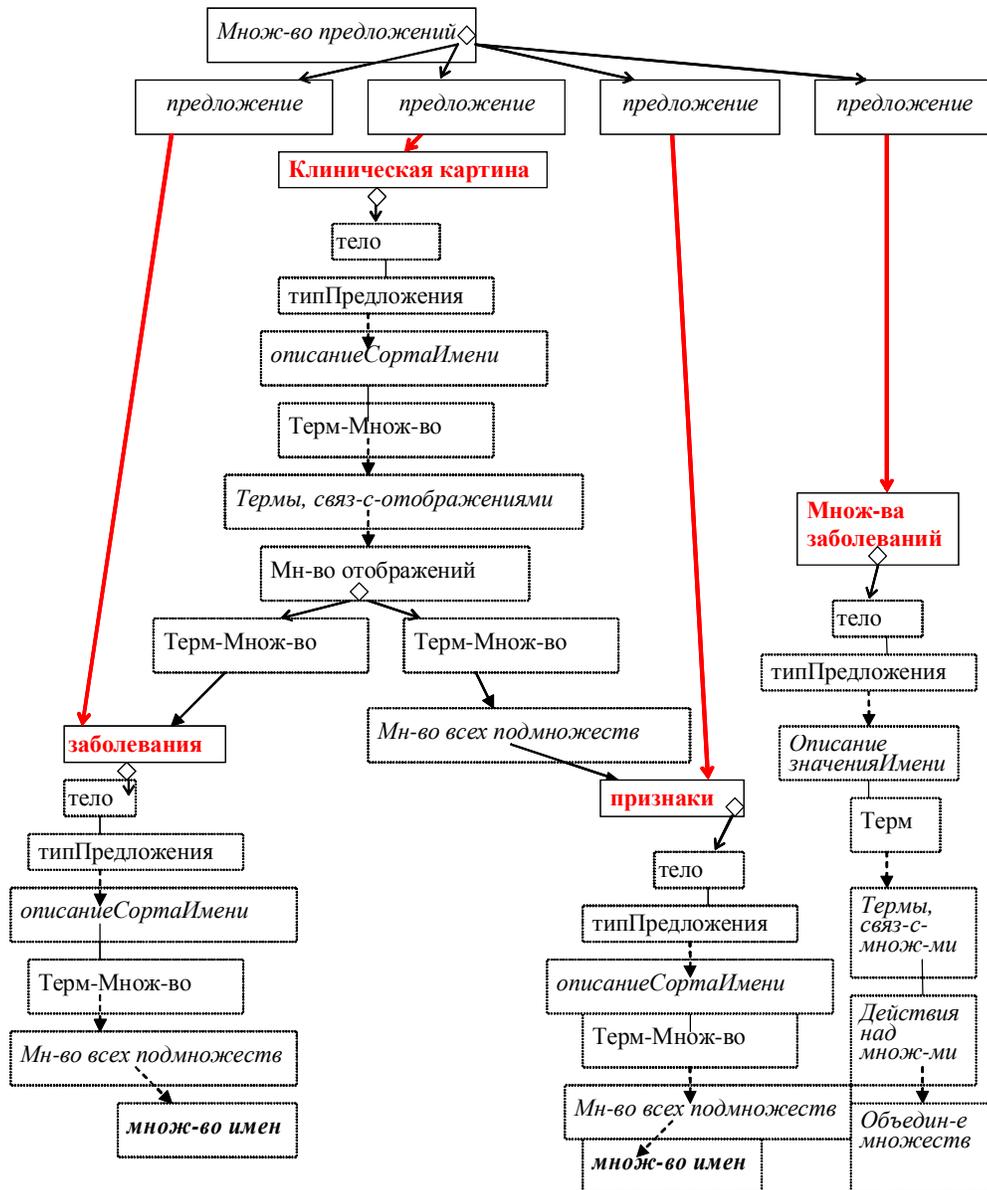


Рис. 2. Фрагмент графа партономических связей сущностей теории 2.

О методе построения *графа теоретико-множественных связей*: терминам хранимой теории (узлам семантической сети) соответствуют вершины; каждому представлению-разбиению термина как *множества терминов* соответствует совокупность дуг «включает элемент», представлению термина в виде *альтернативных терминов* соответствует совокупность дуг «один из», описанию термина как *объекта ИРУО* соответствует дуга «есть».

Модель позволяет определить множество структурных свойств, определения которых представлены в [4, 5]. Рассмотрим несколько примеров свойств, существенных для анализа хранимых в банке ресурсов.

С трудоемкостью проверки определений понятий (в частности после

внесения изменений в те определения понятий, от которых зависят многие другие) связаны такие структурные свойства:

Размер графа теоретико-множественных связей.

Число вершин в графе «теоретико-множественных» связей.

Область значений: неотрицательное целое число.

Длина пути в графе «теоретико-множественных» связей.

Максимальная длина пути от корневой вершины к листовой в графе «теоретико-множественных» связей.

Область значений: неотрицательное целое число.

«Множественность множеств».

Максимальное число дуг «включает элемент» на пути от корневой вершины к листовой в графе теоретико-множественных» связей.

Область значений: неотрицательное целое число.

Наличие циклической теоретико-множественной связи.

Наличие цепочек дуг в графе теоретико-множественных» связей, начинающихся и заканчивающихся одной и той же вершиной.

Область значений: одно из значений {да, нет}.

С корректностью представления совокупности понятий связаны такие структурные свойства.

Число потенциальных «простых» множеств.

Число вершин в графе теоретико-множественных» связей, на пути от которых к любой «листовой» вершине имеется только одна дуга вида «включает элемент».

Область значений: неотрицательное целое число.

Число понятий-множеств.

Число вершин, из которых выходит дуга «включает элемент» в графе теоретико-множественных» связей.

Область значений: неотрицательное целое число.

Число «корневых» множеств.

Число вершин в графе теоретико-множественных» связей, в которые не входят дуги.

Область значений: неотрицательное целое число.

Наличие вершин – элементов нескольких множеств

Наличие вершины, в которую входит более одной дуги «включает элемент» в графе теоретико-множественных» связей

связей.

Область значений: {да, нет}.

Пример 3. В табл. 3 приведены значения свойств структуры теоретико-множественных связей, существенные для фрагмента теории 1 (рис. 3).

Таблица 3

Название свойства	Значение свойства
Число «корневых» множеств	2
Размер графа теоретико-множественных связей	15
Число понятий-множеств	3
Число потенциальных «простых» множеств	2
Наличие вершин - элементов нескольких множеств	Да
Наличие циклической теоретико-множественной связи	да
«Множественность множеств»	2



Рис. 3. Фрагмент графа теоретико-множественных связей сущностей теории 1.

Значения этих свойств дают дополнительную информацию о потенциальной сложности образования терминов как множеств других терминов, часть из которых также являются множествами. С другой стороны, это дает гибкость при построении других теорий, которые будут использовать ее как входную.

Заключение

В работе выделены наборы свойств, существенные для анализа хранимых в банке ресурсов. В частности показаны свойства структуры стандартных связей, характеризующие сложность представления терминов

теории и зависимость онтологий и теорий от базовых ресурсов, на основе которых они построены. Анализ свойств теорий, представляемых такими графами, позволяет получить информацию о полноте представления терминов области знаний в онтологии/теории, о группировании и/или классификации терминов. Использование подхода позволяет получать объективную информацию о свойствах хранимых теорий, что позволяет автоматизировать процесс оценивания этих свойств. Таким образом, показано, что единый подход к определению и оцениванию структурных свойств применим к исследованию, анализу и оцениванию онтологий и других информационно-ресурсов, хранимых в банке знаний.

Перспективными направлениями исследований онтологий и теорий банка с использованием единого подхода являются также: 1) выявление ограничений и возможностей языков представления онтологий по отношению к адекватному моделированию разных областей знаний; 2) выявление реальной использованности языковых конструкций при написании теорий.

ЛИТЕРАТУРА

1. *Gangemi A., Catenacci C., Ciaramita M., Lehmann J.* Ontology evaluation and validation. An integrated formal model for the quality diagnostic task // Technical Report, ISTC-CNR. – Trento, Italy, 2005.
2. *Hartmann J., Sure Y., Giboin A., et al.* Methods for ontology evaluation // Knowledge Web Deliverable D1.2.3. – University of Karlsruhe, 2005.
3. *Brank J., Grobelnik M., Mladenic D.* A survey of ontology evaluation techniques. // Proc. Conf. on Data Mining and Data Warehouses. – Ljubljana, Slovenia, 2005.
4. *Клещев А.С., Шалфеева Е.А.* Принципы организации каталога свойств онтологий // Научно-техническая информация. Сер. 2. – 2007. – № 6. – С.7-15.
5. *Шалфеева Е.А.* Каталог структурных свойств онтологий. Свойства структуры стандартных связей. Препринт. – Владивосток: ИАПУ ДВО РАН, 2007.
6. *Орлов В.А., Клещев А.С.* Многоцелевой банк знаний. Архитектура ИРУО. Владивосток: ИАПУ ДВО РАН, 2003. http://www.iacp.dvo.ru/is/publications/186_4.pdf.

Статья представлена к публикации членом редколлегии А.С. Клещевым.