



УДК 001.53

АНАЛИЗ ДИНАМИКИ ИЗМЕНЕНИЙ НАУЧНОЙ ОБЛАСТИ МЕТОДАМИ ОНТОЛОГИЧЕСКОГО ИНЖИНИРИНГА

Лещева И.А.* , Лещев Д.В.**

* Санкт-Петербургский Государственный Университет,
г. Санкт-Петербург, Россия

leshcheva@gsom.pu.ru

** Центр Перспективных Исследований Санкт-Петербургского Государственного
Политехнического Университета

г. Санкт-Петербург, Россия

dimvovich@gmail.com

Многие современные предметные области подвержены быстрым изменениям. Не смотря на то, что онтологии являются статическим срезом представлений о предметной области, серия онтологий может быть использована для анализа динамики предметной области. В работе описан общий алгоритм построения серии онтологий, отражающих изменения во времени, для последующего анализа. Алгоритм апробирован на предметной области знаний об углеродных нанокластерах, отраженных в тезисах докладов конференции IWFAC.

Ключевые слова: онтологии; серия онтологий.

Введение

Развитие предметных областей приводит к тому, что онтологии, отражающие знания о них, также изменяются. В настоящее время существует ряд работ, посвященных проблеме изменения онтологий во времени. Их авторы сосредоточены на том, чтобы при внесении изменений/дополнений онтология оставалась непротиворечивой. В таких работах само изменение онтологии является проблемой, которую надо решить, например, с помощью специального инструментария [Загорулько, 2007; Farquhar, 1997; Karp, 1999].

С другой стороны, развивается направление, посвященное разработке временных (темпоральных) онтологий, например, [De Beule, 2004]. Основная идея — сделать онтологии пригодными для моделирования динамики изменения предметных областей.

В данной работе именно изменения онтологии во времени (динамика) являются предметом анализа, но без прямого построения темпоральных функций. С помощью подобного изучения динамики онтологий можно отслеживать развитие знаний в любой предметной области на базе статического онтологического подхода.

1. Процесс построения онтологий

1.1. Общий алгоритм

Общий алгоритм, позволяющий построить онтологию для будущего анализа их динамики можно разбить на три шага.

Шаг 1. Сначала строится базовая онтология, покрывающая все основные концепты предметной области [Gavrilova, 2007]. Если необходимо, процесс построения базовой онтологии может быть итеративной.

Шаг 2. Полученная на предыдущем шаге онтология расширяется и уточняется экспертами с учетом изменений знаний во времени. В результате нескольких итераций получается всеобъемлющая онтология, охватывающая все более-менее существенные концепты предметной области. Данная онтология может служить классификатором или рубрикатором для всех возможных тем в данной области знаний.

Шаг 3. Строятся проекции общей онтологии на некоторые заранее определенные моменты времени. По факту, все устаревшие или еще не появившиеся концепты-листья и даже целые ветви исключаются. Также на данном этапе возможно сопоставление каждому концепту некоторого количественного

показателя, например, число публикаций по данному вопросу. Именно этот показатель и становится в будущем основой анализа темпоральных изменений.

1.2. Применение алгоритма

В качестве примера предметной области для апробации алгоритма была взята конференция «Фуллерены и атомные кластеры» ("Fullerenes and Atomic Clusters", IWFAC), которая в 2011 году была преобразована в конференцию «Продвинутые углеродные наноструктуры» (International Conference "Advanced Carbon Nanostructures" (ACN'2011)) [ACN, 2013] в связи с сокращением числа выступлений, связанных с неуглеродными кластерами. Конференции являются удобным средством для отражения концептов определенной небольшой предметной области: они меньше подвержены отбору в отличие от статей в научных журналах и позволяют легко регулировать глубину строящейся онтологии.

Онтологии строились на материале сборников тезисов конференции (биеннале с 1993 по 2009, всего 9 конференций, 300-400 докладов на каждой конференции). Применительно к данной задаче алгоритм упростился:

Шаг 1. Была построена всеобъемлющая онтология, покрывающая тематику докладов самой «богатой» по темам конференции IWFAC'07. Верхние уровни ветви «Углеродные материалы» приведены на рисунке 1. Реальная онтология весьма велика, и помещать ее целиком в статью не имеет смысла. Самым верхним уровнем концептов стали «Углеродные материалы» и «Неуглеродные материалы»

Шаг 2. Полученная на предыдущем шаге онтология была расширена экспертом за счет потенциально-возможных областей исследования, не освещенных в докладах конференции 2007 года. Таким образом полученная на 2 шаге онтология послужила рубрикаторм для тем докладов.

Шаг 3. Доклады каждого года были распределены по темам в соответствии с полученным на 2 шаге рубрикаторм. Для всех узлов каждой из 9 полученных онтологий, помимо названий концептов, были рассчитаны веса, показывающие, сколько докладов было сделано по соответствующей теме.

Всего было выделено 8 уровней иерархии. Второй уровень иерархии определялся наличием трансляционной симметрии материала: кристаллы, кластеры (молекулы), полимеры и аморфные вещества. Следующее деление зависело от развитости концепта второго уровня. Например, кластеры были разделены на фуллерены, нанотрубки, наноалмазы, онионы и остальные. Четвертый уровень — это уровень модификации

чисто углеродной основы: немодифицированные, производные, комплексы, эндоэдральные и допированные вещества. Следующий уровень определял вид модификации. И последний уровень иерархии содержал собственно темы тезисов. Если тезисы могли быть отнесены к нескольким разным первичным концептам, то доклад становился листом более общего родительского концепта. Таким образом, у каждого концепта появился вес, равный количеству докладов, отнесенных к данному концепту.

При классификации каждого доклада учитывалась только ключевая идея тезисов. Построенная онтология достаточно субъективна, и тезисы можно классифицировать и другими, отличными от предложенного образцами. Это отличает данный способ построения онтологии/таксономии от машинных способов классификации, которые используют вхождения всех употребляемых слов формально. Например, в большинстве докладов во вводной части говорится о фуллеренах, как о символе новых форм углерода, хотя сама работа может быть посвящена углеродным нанотрубкам, полимерам или вообще неуглеродным материалам.

2. Анализ динамики предметной области

Построенная последовательность онтологий была проанализирована с точки зрения изменений во времени.

Типичный сценарий развития темы состоит из 3 основных стадий: рождения, роста и угасания. Для примера на рисунке 2 приведена диаграмма числа докладов по теме «Органические производные фуллеренов» и ее подтемам за 10 лет. Для удобства визуализации использована логарифмическая цветовая шкала. Из рисунка, например, видно, что максимум исследований галогенсодержащих органических производных фуллеренов приходится на 2007 г.

Если рассматривать две связанные темы «Галогенсодержащие органические производные фуллеренов» и «Галогениды фуллеренов», то видно, что рост первой темы сопровождается уменьшением числа докладов по второй. Можно предположить, что исследователи заканчивают работу над исчерпавшей себя темой и переходят в близкую. Подобный феномен наблюдается во многих парах.

На рисунке 1 можно также заметить, что концепты зрелых областей исследования содержат 3 дочерние ветви: синтез, свойства и применения. В молодых областях такого разделения нельзя провести. Согласно этому выводу, к зрелым областям можно отнести только две темы: немодифицированные фуллерены и немодифицированные нанотрубки.

Carbon materials

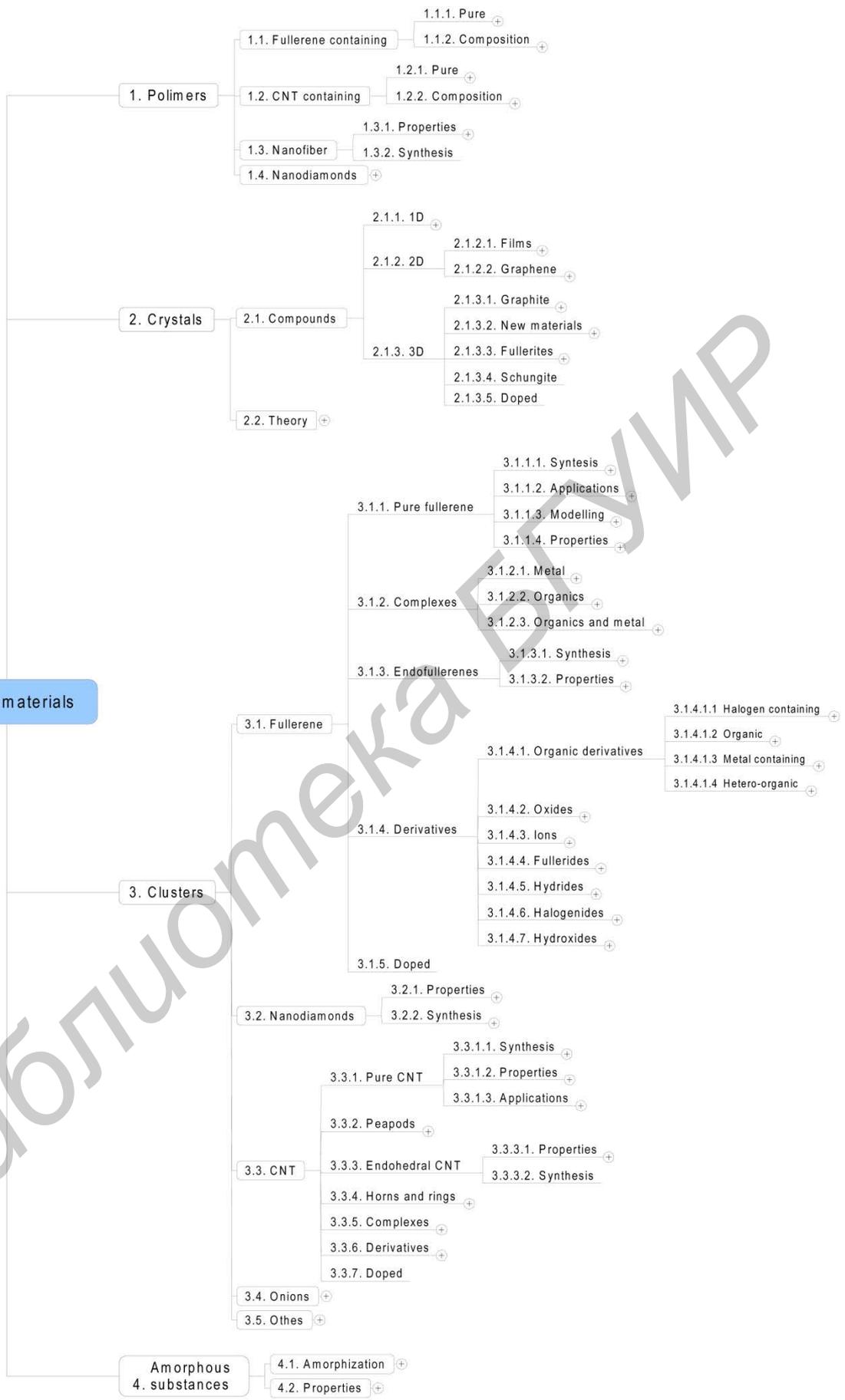


Рисунок 1 — Верхние уровни ветви «Углеродные материалы» общей онтологии

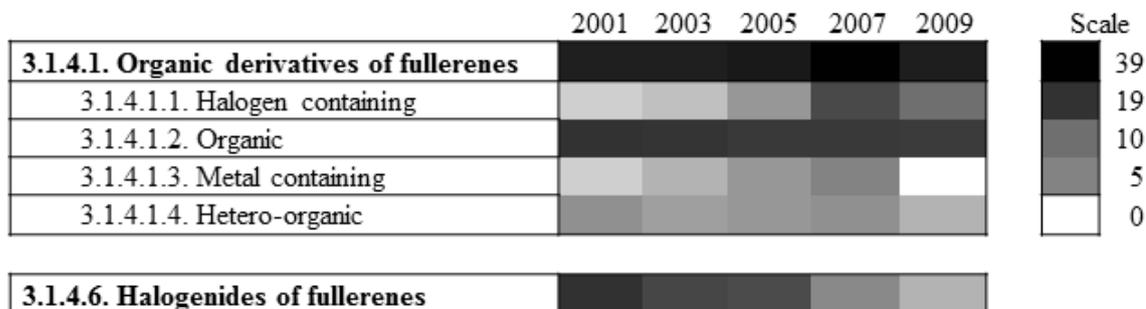


Рисунок 2.— Эволюция популярности двух тем

Заключение

В работе предложен подход использования онтологий для анализа динамики предметных областей, формализованный в виде общего алгоритма. Алгоритм апробирован на предметной области, охватывающей темы докладов одной из конференций, посвященной углеродным нанокластерам.

Были отмечены следующие закономерности:

- Типичный сценарий развития темы состоит из 3 основных стадий: рождения, роста и угасания.
- «Молодые» темы являются листьями онтологии, в то время как развитые области содержат разветвленную структуру.
- Некоторые темы связаны между собой: потеря интереса к одной из областей соответствует росту исследований в другой. Можно предположить, что развитие второй темы сопровождается оттоком в нее исследователей из области.

Таким образом, в результате анализа изменения онтологий во времени можно проследить эволюцию отдельных тем, выявить тренды развития, оценить перспективность того или иного направления для исследований или инвестиций и даже предсказать, что будет происходить в ближайшем будущем.

Библиографический список

- [Загорюлько, 2007] Загорюлько Ю.А., Боровикова О.И. Технология построения онтологий для порталов научных знаний/ Вестник Новосибирского государственного университета. Серия: Информационные технологии, 2007, Т.5, № 5, сс.42-52.
- [ACN, 2013] 11th International Conference Advanced Carbon NanoStructures Home Page <http://www.ioffe.ru/ACNS/index.html> (дата обращения: 30.11.13).
- [De Beule, 2004] De Beule, J. Creating Temporal Categories for an Ontology of Time. In Rineke Verbrugge and Niels Taatgen and Lambert Schomaker, editors, BNAIC-04 [Электронный ресурс] / De Beule, J. Creating. – P. 107-114. – Режим доступа: <http://arti.vub.ac.be/~joachim/Bnaic04.pdf>
- [Farquhar, 1997] Farquhar A., Fikes R., Rice J. The Ontolingua Server: a tool for collaborative ontology construction/ International Journal of Human-Computer Studies, 1997, V. 46, I.6, pp. 707-727.
- [Gavrilova, 2007] Gavrilova, T., Ontological Engineering for Practical Knowledge Work, Lecture Notes in Artificial Intelligence, 2007, 4693, pp. 1154-1162.
- [Karp, 1999] Karp P., Chaudhri V., Paley S. A Collaborative Environment for Authoring Large Knowledge Bases/ Journal of Intelligent Information Systems. 1999, V.13, I.3, pp.155-194.

ANALYSIS OF TRENDS IN SCIENTIFIC DOMAINS BY ONTOLOGICAL ENGINEERING

Leshcheva I.A. *, Leshchev D.V. **

* Graduate School of Management, St. Petersburg University, St.Petersburg, Russia

leshcheva@gsom.pu.ru

**Center for Advanced Studies, St. Petersburg State Polytechnic University, St. Petersburg, Russia

dimvovich@gmail.com

In this work a general algorithm for constructing a scientific domain ontologies to analyze the scientific domain dynamics is proposed. The algorithm is tested on an analysis of conference abstracts for 20 years.

Introduction

The subject of this work is applying ontology engineering approach to analyzing of trends in scientific domains. By constructing series of static ontologies at different moments of time the dynamics of knowledge in a subject area can be traced.

Creating of ontologies series

A three steps algorithm is proposed. On the first step, a basic ontology covering all main concepts of the domain is constructed. On the second step, the obtained ontology is expanded by experts. These steps can be iterative. On the third step, a projection of the received ontology on a particular point of time is built.

Domain trends analysis

The described algorithm was used to analyze dynamic changes in carbon nanoclusters domain using abstracts of International Workshop “Fullerenes and Atomic Clusters” (IWFAC) [ACN, 2013].

Conclusion

The proposed approach can be used to define points of domains growth and stagnation, relationship between subdomains and there maturity.