

Andrej Poleev. Universal Metadata Standard.

Abstract. The basis of consciousness is an association of notions, the neuronal network. Similarly, the creation of a next generation internet (semantic web) is impossible without attributes, allowing the semantic association of documents and their integration into information context. To achieve these goals, the Universal Metadata Standard (ums) may be an ultimate tool, which could serve as a basis for documentography, and is functionally required for interpretation of documents by the automatic operating systems.

Key words: document, metadata, classification, identification, association, documentography, metagraphy, standard, metabase.

Резюме. Основой сознания является ассоциативная связь понятий, согласованная работа элементов нейрональной сети. Аналогично этому, создание интернета нового поколения (semantic web) невозможно без атрибутов, позволяющих осуществлять семантическую связь документов и интеграцию их в информационный контекст. Для реализации этих целей предлагается ввести Универсальный Стандарт Метаданных (universal metadata standard, ums), который мог бы служить основой документографии (documentography), функционально необходимой для интерпретации документов в автоматических операционных системах.

Ключевые слова: документ, метаданные, классификация, идентификация, ассоциация, документография, метаграфия, стандарт, база метаданных.

Цель научного познания – объять необъятное. Невозможность достижения этой цели очевидна, однако если принимать её за максимум научно-познавательной деятельности и путеводную звезду в поисках истины, удовлетворение стремления узнать больше и расширить индивидуальный круг знаний представляется вполне разумным и оправданным мотивом любого человека. В сущности, значительную часть времени люди заняты организацией информационного потока, непрерывно поступающего в их мозг через органы чувств и рецепторы как из тела так и извне. Не только их благополучие, но и шансы на выживание, определяются тем, насколько эффективно происходит такое упорядочивание информации, в результате чего сырьё нервных импульсов превращается в достоверное знание.

Появление и развитие сознания связано с совершенствованием средств коммуникации, основанием чего является знаковая передача информации, язык. Непрерывное совершенствование техники коммуникации, преодоление семантических барьеров методом проб и ошибок, привело к возникновению стандартов передачи и восприятия информации, примером чего является книгопечатание (полиграфия). Продолав значительный путь, полиграфическая техника привела к становлению микроэлектроники, которая не только улучшила качество и расширила область достоверного знания, но и ознаменовала собой возможность злонамеренного манипулирования сознанием, поскольку из поля зрения читателей и зрителей, т.е. реципиентов информации, стали исчезать сферы производства и распространения знания, а также те аспекты документов, которые недоступны непосредственному человеческому восприятию, однако могут или должны быть восприняты обрабатывающими информацию машинами (компьютерами). Восполнить возникший пробел восприятия представляется важной задачей информатики.

Рассмотрим пример того, как происходит организация знания. В центре внимания учёного сообщества находится проблема накопления, верификации и систематизации знаний, оформляемых в виде научных публикаций. Однако появлению всякой публикации предшествует значительная деятельность, как правило сокрытая от публики. Черновой вариант научных статей – лабораторный журнал – это не что иное как сборник протоколов о запланированных экспериментах и их результатах. Однако в идеальном случае он должен протоколировать всё, относящееся к проводимой научной работе и отражать всё, что происходит в лаборатории в хронологическом порядке, начиная от целеполагания, гипотезы, экспериментальной проверки, выводов, впечатления об увиденном и

услышанном. В формальном плане, лабораторный журнал должен описывать документы различного формата: фотографии, тексты протоколов, тексты публикаций, видеозаписи конференций (lab meetings), указания на источники в интернете и т.д. Все эти документы должны быть связаны между собой, снабжены комментариями, и доступны для просмотра и каталогизации. Например, в хронологическом порядке друг за другом могут следовать эксперименты или мысли, относящиеся к различным темам: теоретическое исследование определённого вопроса и сбор соответствующей информации; написание статьи или книги на основании уже завершённой работы; планирование тематически разнородных экспериментов. В связи с этим эта тематическая разнородность должна быть отражена в списках тем, а также в возможности экстрагировать однородную (родственную) информацию посредством указателей (thematic tags) и ссылок (location tags).

Компьютер MacBook, которым я пользуюсь, предоставляет возможность накапливать и тематически объединять разнородные документы. Однако для их описания, воспроизведения или визуализации необходимо дополнительное программное обеспечение. File Maker только частично удовлетворяет потребности систематизации и описания: на данном этапе отсутствует приемлемая панель обзора и возможность открывать и использовать документы внутри данной программы, не прибегая к дополнительным программам. Все эти дополнительные программы в идеальном случае должны быть встроены в качестве опций, а не разбросаны по разным местам: web editor, web browser, photoshop, file maker, pdf reader, video or photo visualiser, text editor и т.д..

В связи с многообразием документальной основы научного сознания и познания, проблема документации и систематизации данных приобретает первостепенное значение. Обычно документы классифицируют по алфавиту, по дате, по теме, по проекту, по формату, по местонахождению (local folder, internet address). Для их идентификации служит дата, порядковый или систематический номер, имя (название). Например, изображения (images) имеют формат (file format) jpg, gif, png, psd; тексты (texts) имеют формат pdf, doc, txt. Формат документа – это его идентификационный признак (identification tag), необходимый для опознания в операционных системах и инициации программ (процессирования). Однако в каждом формате до сих пор отсутствует его систематическое описание, необходимое и достаточное для интеграции и переноса в другие описательные системы (например, при копировании из электронной библиотеки в персональный компьютер). Всякий документ отражает реальные предметы и события, является их описанием, отображает определённые качества. Однако фотография не сохраняет информации о размерах объекта, о его происхождении, истории, цели. Всё это в идеальном случае должно входить в метаинформационное дополнение документа, по крайней мере в виде ссылок. Однако увеличение количества документов и форматов не сопровождается совершенствованием технических возможностей их восприятия и систематизации. Вместо этого происходит разможение описательных систем (doi, ISBN, URN, PURL, ISNI и др.) и псевдонимов (aliasing). Так например, журнальная статья, как правило в форматах html или pdf, в описательной системе NCBI/NLM получает номер (PUBMED ID), добавляется резюме (abstract) с сопряжённым указанием на время публикации, название журнала, имён авторов, языка, ключевых слов. Необходимо однако, чтобы эта описательная метаинформация добавлялась непосредственно в документ в качестве дополнения или расширения, чтобы было возможно упорядочивание документа при перемещении его в другие описательные системы (например, при переводе на другой язык, или при использовании в другой базе данных), а история такого перемещения (например, при копировании из электронной библиотеки) отображалась бы в документе. Для достижения этой цели

следует создать универсальный стандарт для всех типов документов, и договориться о том, какие опции будут присутствовать в каждом формате; как их будут заполнять или модифицировать; что не должно подвергаться изменению. Мне представляются очевидными нижеперечисленные опции метаинформационного описания документов:

имя * (preferably unique name)
формат (format)
дата создания (date)
классификационная система (classification system used)
идентификационный номер (identity number)
язык ** (language)
локализация или место происхождения (position, location)
источник или автор (creator, origin, source)



* Систематическое имя (systematic designation) – это последовательность символов (знаковая секвенция), на основании которой происходит идентификация обозначаемого объекта и установление соответствия между восприятием его посредством органов чувств (сенсорной репрезентацией) и языковой интерпретацией этого восприятия. Систематическое имя должно отличаться качествами, позволяющими отнести его к классу наименований, а также содержать необходимое дополнение, достаточное для однозначной идентификации среди родственных названий, обозначений и имён. Например, в узком круге лиц, достаточным является имя Андрей, в то время как в группе, имеющей в своём составе несколько людей с тем же именем, необходимо указывать родовое имя (фамилию) для различения. В планетарном масштабе, достаточным является указание имени, даты и места рождения для установления идентичности. Систематическое имя для обозначения персон может состоять из двух или трёх имён, последовательности цифр, и географического определителя. Аналогично этому, систематическое имя организаций может содержать название, указание на дату и место основания, адрес, дату завершения деятельности. Ответ на 3 вопроса: Кто или что? Где? и Когда? является достаточным для идентификации и в других случаях.

Понятия каталог, номенклатура, классификация, регистр в значительной мере синонимичны, и обозначают список имён, объединённых в родственные группы, которые в свою очередь также сгруппированы на основании определённых критериев. Порядок группирования может изменяться в зависимости от выбранных критериев. Имена персон можно группировать по алфавиту, на основании даты или места рождения их прообразов. В динамичном пространстве категоризации, систематическое имя остаётся константой, кристаллизационным пунктом, отправной точкой в процессе семантической ассоциации, поиска и установления отношений и взаимосвязей между именами, понятиями, определениями, категориями.

** язык подразумевает знаковые системы естественных языков, имеющих дескриптивный и индикативный характер; языки программирования являются производными естественных языков, и имеют директивный характер алгоритмов, т.е. инструкций для автоматических операторов

Понятно, что инструкции по производству атомного оружия, или документы порнографического характера не могут быть доступны всем кому не лень. Поэтому для ограничения доступа к документам следует ввести градацию доступности.

Если документ будет претерпевать модификации (перенос в другую описательную систему, изменение размера, формата, названия), то первичные метаданные должны сохраняться, а изменения автоматически или мануально записываться: при переименовании добавляться синонимическое имя; в другой описательной системе (системе классификации) добавляться её обозначение и идентификационный номер в этой системе; при транспозиции записываться новый адрес в интернете или географическое соответствие, и т.д.

Для каждого атрибута стандарта УМ следует определить форму опции, дать её определение и формальное описание. Содержание каждой опции должно соответствовать правилам, на основании которых составлялся бы каталог допустимых значений (metabase: catalog of systematic designations). Например, авторство документов должно быть однозначным на основании списка авторов. Происхождение документа должно указываться на основании списка организаций. Указание типа документа (текст, рисунок, фотография, видео, звук), должно сопровождаться описанием (резюме), и типологической атрибутикой, характерной для каждого типа документов. Каждый документ должен содержать перечень объектов или явлений, отображением или описанием которых он является (биологический вид, астрономический объект, персона или группа лиц, организация, научная публикация и т.д.). Классификационная основа такого перечисления в настоящее время существует, (Encyclopedia of Life, International Plant Names Index, Catalogue of astronomical objects, PubMed, ICD и другие), следует использовать её в стандарте УМ.

Что же происходит в реальности? Рассмотрим показательный пример. Экстракция метаданных для документа octology.pdf, имеющего адрес <http://www.enzymes.at/download/octology.pdf>, дала следующий результат:

```
CreateDate = 2011:03:01 16:35:22Z
Title = octology
PageCount = 76
FileSize = 11 MB
Author = Max Madman
MIMEType = application/pdf
PDFVersion = 1.4
FileType = PDF
Creator = Pages
ModifyDate = 2011:03:01 16:35:22Z
PDFVersion (1) = 1.3
Producer = Mac OS X 10.5.2 Quartz PDFContext
```

Очевидна бессмысленность такого описания: указание на формат (pdf) присутствует 6 раз; кто создатель и автор документа – неясно; время создания и модификации документа совпадают и ничего не сообщают о времени его появления на свет божий. Пожалуй только указание на количество

страниц и размер документа является осмысленным. Метаинформация включённых в текст иллюстраций (если таковая имела) полностью утрачена в формате pdf. Публикация документа на портале Researchgate.net сопровождается указанием адреса DOI: details/Octology. Что это значит, неясно, поскольку проверка этого адреса в описательной системе DOI не приводит ни к какому результату. Хотя журнал, в котором осуществилась публикация данного документа, включён в базу данных NCBI/NLM, однако информация об этом документе до сих пор не появилась в электронной библиотеке PUBMED. Никакой возможности сделать это вручную нет: всё предоставлено программному обеспечению и автоматам, которые однако не в состоянии осуществить возложенную на них функцию. Сбой системы проявляется ещё и в том, что информация о документе с названием Octology, отсутствует в PUBMED/NCBI/NLM, но присутствует в связанной с ней описательной системе OCLC/WorldCat.

Ещё абсурднее выглядят метаданные документа, наугад выбранного из библиотеки PUBMED: Palesch D, Sieńczyk M, Oleksyszyn J, Reich M, Wiczerzak E, Boehm BO, Burster T. Was the serine protease cathepsin G discovered by S. G. Hedin in 1903 in bovine spleen? Acta Biochim Pol. 2011 Mar 7, PMID: 21383996 (см. Приложение).

Подводя итоги, следует предложить программистам, терминологам, ISO и индустрии знаний разработать логически выверенную систему метаинформационного обеспечения для всеобщего пользования, чтобы производство документов не превратилось в самоцель, а приобрело бы надёжную основу, позволяющую на новом этапе социального и технического развития эффективно усваивать и организовывать знания. Параллельно, следовало бы дополнить существующие программы модулем, позволяющим визуализировать и редактировать метаданные, а также ввести в употребление универсальные программы для всех типов документов (metadata editors).

Более общей тематической идеей данной статьи является создание комплекса семантических стандартов, частью которых может стать стандарт УМ. Никола Тесла, заложивший технические основы для создания интернета на рубеже 19 и 20 веков, руководствовался мечтой об упразднении границ, препятствующих общению и познанию. Сегодня интернет, став реальностью, сам создаёт виртуальную реальность, на основании которой конструируется действительность, сознание и общество. Необходимо указать на связанные с этим опасности. Так например, смысловое содержание одного из центральных понятий интернета 3 поколения – онтология***, умышленно искажается в идеологических целях: Онтологиями в бизнесе стали обозначать логические схемы, разработанные для манипуляции сознания, вбивания заранее заданных стереотипов мышления, пропаганды групповых интересов. Написанные на непонятном широкой публике искусственном языке, онтологические схемы призваны осуществлять скрытый контроль над обществом узким кругом лиц, определяющих правила их написания. В связи с этим, семантический интернет может стать инструментом тоталитарного управления, имеющего глобальный характер. Понятно, что захват власти может осуществиться конспиративно, а сам тоталитарный процесс в таком случае будет вынесен за рамки юридического регулирования. Чтобы исключить злонамеренное использование технологии интернета, необходимо своевременно принять упреждающие меры. Предлагаемые в данной статье универсальные стандарты позволят избежать данного развития и сделать семантический интернет более осмысленным, реалистичным и доступным для регулирования широким кругом его пользователей.

*** Поскольку бытие (сущность) объектов проявляется в действии, то комплексное описание взаимодействия в данном множестве объектов даёт наиболее полное представление об изучаемой области. Онтологическая схема – это формализованное описание связей и взаимодействия между объектами в определённом множестве объектов. Примером применения онтологических схем может служить научная область, включающая всю совокупность изучаемых явлений и объектов, методов изучения и описания, гипотез и теорий. Другой пример: производственное предприятие, являющееся совокупностью оборудования (средств производства), технологического описания производства (методов производства), правил поведения персонала (инструкций управления предприятием) и других условий его функционирования.

В центре онтологической схемы находится описание объектов, включающее наименование или адресацию и установление свойственных им атрибутов (качеств и свойств проявления). Всякое описание основывается на систематизации, позволяющем отнести описываемый объект к группе объектов данной онтологической схемы. При этом атрибуты объектов могут приобретать более общий характер систематических категорий, на основании которых всё множество объектов распределяется на подгруппы. Например, во множестве предметов, некоторые из них могут быть шаровидной формы, отличаться по цвету и т.д. Таким образом, различение объектов происходит путём систематизации на основании индивидуальных признаков, а категоризация является рекурсивной операцией, выделяющей необходимые и достаточные признаки объектов, на основании которых осуществляется их систематизация и распределение внутри данного множества объектов.

Однако онтологические схемы могут не только описывать данность, но и активно влиять на объекты, определять их поведенческий модус посредством установления правил взаимодействия. Субъективный фактор онтологических схем наглядно проявляется в государственном управлении, основывающемся на неполном, искажённом, или неадекватном описании объектов, т.е. людей, социальных групп и их взаимоотношений, а также исключаящем из рассмотрения онтологические схемы более общего характера (экология, биосфера, космология, философия). Неудивительно, что люди в таких онтологиях до сих пор рассматриваются как расходный материал, с которым можно обращаться как с неодушевлёнными предметами или домашним скотом.

Неосведомлённый читатель может восполнить пробел знаний, ознакомившись со следующими источниками:

Семантический интернет.

Michael K. Bergman. A Timeline of Information History.

Handbook of Metadata, Semantics and Ontologies, 2012, World Scientific Publishing.

Tom Heath and Christian Bizer (2011) Linked Data: Evolving the Web into a Global Data Space (1st edition).

Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool. ISBN: 9781608454310.

Бионтологии.

Онтология музыки.

Онтологии для электронного правительства.

Конструирование онтологий.

Метаданные документов.

Список форматов документов.

Adobe XMP.

Стандарты метаданных.

eXtended MetaData Registry (XMDR) Project.

Introduction to Metadata, 2008, by Tony Gill, Anne J. Gilliland, Maureen Whalen, and Mary S. Woodley, Murtha Baca (Ed.)

RIP Keywords Meta Tag.

Библиография о метаданных (метаграфия)

Steven J. Miller. Metadata and Cataloging Online Resources, 2010.

Greenberg, Jane. Metadata and Digital Information. In Marcia J. Bates, Mary Niles Maack, Miriam Drake eds. Encyclopedia of Library and Information Science, 2009, New York: Marcel Dekker, Inc.

Метаданные в искусстве, литературе и философии

Mark Amerika META/DATA A Digital Poetics, 2007, MIT Press.

Metaexhibition.

The Handbook of Organization Theory: Meta-theoretical Perspectives by Haridimos Tsoukas, Christian Knudsen. O.r.d Un.ty Press, 2003, ISBN: 0199258325.

Metadata Symposium, Academy of Motion Picture Arts and Sciences.

Лабораторный журнал.

Amber Dance. How to choose your lab's next electronic lab notebook. The Scientist, 2010, vol. 24, 5, p. 71.

Технические средства организации научной литературы

EndNote.

Mekentosj Papers.

Приложение.

Экстракт метаданных упомянутого в тексте статьи документа из библиотеки PUBMED; экстракция произведена с помощью <http://www.serversniff.net/file-info.php>.

1. <http://www.ncbi.nlm.nih.gov/pubmed/21383996>

FileType(guessed) = XML

title - Was the serine protease cathepsin G discovered by ... [Acta Biochim Pol. 2011] - PubMed result

keywords - PubMed, National Center for Biotechnology Information, NCBI, United States National Library of Medicine, NLM, MEDLINE, Medical Journals, pub med, Entrez, Journal Articles, Citation search

description - PubMed is a service of the U.S. National Library of Medicine that includes over 19 million citations from MEDLINE and other life science journals for biomedical articles back to the 1950s. PubMed includes links to full text articles and other related resources.

author - pubmeddev

ncbi_stat = false

ncbi_phid = CE8875A4D78C969100000000001DA980

ncbi_pdid = abstract

Keywords = PubMed, National Center for Biotechnology Information, NCBI, United States National Library of Medicine, NLM, MEDLINE, Medical Journals, pub med, Entrez, Journal Articles, Citation search

Title = Was the serine protease cathepsin G discovered by ... [Acta Biochim Pol. 2011] - PubMed result

ncbi_pagesize = 20

ncbi_filter = all

FileSize = 92 kB

ncbi_format = html

ncbi_uidlist = 21383996

Author = pubmeddev

ncbi_report = abstract

ncbi_resultcount = 1

MIMEMType = text/html

ncbi_app = entrez

ncbi_hitstat = true

FileType = HTML

ncbi_pageno = 1

ncbi_db = pubmed

ncbi_sortorder = default

ncbi_sessionid = CE8875A4D78C9AD1_0029SID

Robots = index,nofollow,noarchive

ncbi_op = retrieve

Description = PubMed is a service of the U.S. National Library of Medicine that includes over 19 million citations from MEDLINE and other life science journals for biomedical articles back to the 1950s. PubMed includes links to full text articles and other related resources.

2. <http://www.actabp.pl/pdf/Preprint/20112013.pdf>

FileType(guessed) = PDF document, version 1.4

format - PDF 1.4

mimetype - application/pdf

MetadataDate = 2011:03:06 19:04:38+01:00

ManifestReferenceDocumentID (7) = xmp.did:E981B13A4B44E0118B25D8B18DDAEA80

ManifestReferenceInstanceID (6) = xmp.iid:E881B13A4B44E0118B25D8B18DDAEA80

ManifestPlacedResolutionUnit (8) = Inches

ManifestLinkForm (4) = ReferenceStream

DocumentID = xmp.did:36A96DDD4444E011BA44C4547900889D

ManifestLinkForm (7) = ReferenceStream

ThumbnailHeight = 256

ManifestLinkForm (9) = ReferenceStream

ManifestReferenceDocumentID (3) = xmp.did:ED81B13A4B44E0118B25D8B18DDAEA80

ManifestPlacedResolutionUnit = Inches

MIMEType = application/pdf

FileType = PDF

ManifestReferenceDocumentID (5) = xmp.did:2F33EFA24A44E0118B25D8B18DDAEA80

ManifestPlacedYResolution (8) = 300.00

ManifestReferenceInstanceID (5) = xmp.iid:3333EFA24A44E0118B25D8B18DDAEA80

ManifestPlacedResolutionUnit (3) = Inches

DerivedFromOriginalDocumentID = adobe:docid:indd:fac21e1b-e9a1-11db-bb63-f87ecfa109dc

ManifestPlacedResolutionUnit (5) = Inches

ManifestPlacedXResolution (7) = 300.00

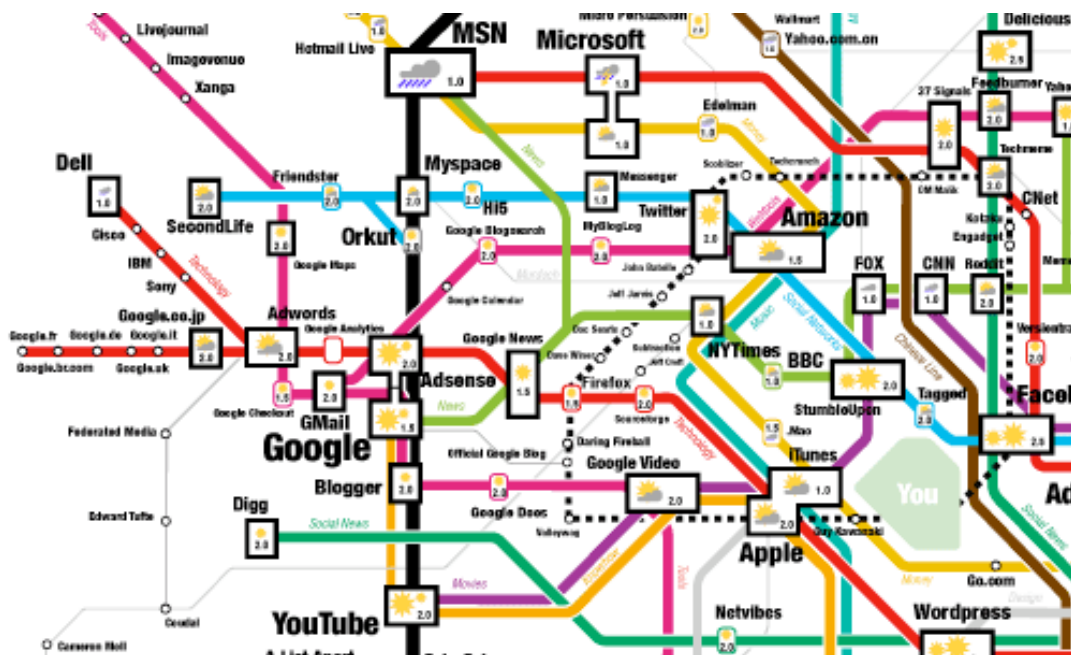
ManifestLinkForm = ReferenceStream

ManifestReferenceDocumentID (8) = xmp.did:EB81B13A4B44E0118B25D8B18DDAEA80

HistoryWhen = 2010:01:08 10:37:37+01:00, 2010:01:08 10:37:37+01:00, 2010:01:08 10:38:47+01:00,
2010:01:09 20:37:04+01:00, 2010:02:15 17:38:17+01:00, 2010:02:15 17:38:18+01:00, 2010:02:15
17:54:28+01:00, 2010:02:15 18:09:18+01:00, 2010:02:15 18:10:28+01:00, 2010:02:15 18:28:53+01:00,
2010:02:15 18:35:15+01:00, 2010:02:15 18:54:59+01:00, 2010:02:15 18:57:27+01:00, 2010:02:15
23:31:26+01:00, 2010:02:15 23:34:43+01:00, 2010:02:15 23:43:09+01:00, 2010:02:15 23:47:48+01:00,
2010:02:15 23:51:01+01:00, 2010:02:15 23:52:57+01:00, 2010:02:15 23:58:11+01:00, 2010:02:16
00:00:58+01:00, 2010:02:16 00:03:51+01:00, 2010:02:16 10:16:48+01:00, 2010:02:16 10:26:27+01:00,
2010:02:16 11:03:51+01:00, 2010:02:16 11:28:37+01:00, 2010:02:16 15:07:05+01:00, 2010:02:16
15:09:02+01:00, 2010:02:16 15:13:01+01:00, 2010:02:16 15:21:18+01:00, 2010:02:16 15:35:34+01:00,
2010:02:16 15:44:47+01:00, 2010:02:16 18:18:01+01:00, 2010:02:16 18:58:14+01:00, 2010:02:16
18:58:36+01:00, 2010:02:16 19:19:28+01:00, 2010:02:16 19:48:35+01:00, 2010:02:16 19:57:42+01:00,
2010:02:16 20:25:24+01:00, 2010:02:16 22:38:42+01:00, 2010:02:24 12:57:22+01:00, 2010:02:24
12:57:22+01:00, 2010:02:24 12:59:57+01:00, 2010:02:24 13:06:53+01:00, 2010:02:24 16:25:01+01:00,
2010:02:24 16:27:04+01:00, 2010:02:24 19:56:39+01:00, 2010:02:24 19:56:50+01:00, 2010:02:24
19:56:50+01:00, 2010:02:28 19:45:50+01:00, 2010:02:28 19:45:50+01:00, 2010:02:28 19:46:03+01:00,
2010:02:28 19:46:03+01:00, 2010:02:28 20:56:51+01:00, 2010:02:28 20:57:03+01:00, 2010:02:28
20:57:03+01:00, 2010:02:28 20:58:25+01:00, 2010:03:08 10:02:08+01:00, 2010:03:08 10:02:08+01:00,
2010:03:09 21:12:26+01:00, 2010:03:09 21:12:26+01:00, 2010:03:13 00:20:22+01:00, 2010:03:13

xmp.iid:585E8AC43B21DF118E14AB5AA9CB3F6F, xmp.iid:595E8AC43B21DF118E14AB5AA9CB3F6F,
xmp.iid:5A5E8AC43B21DF118E14AB5AA9CB3F6F, xmp.iid:5B5E8AC43B21DF118E14AB5AA9CB3F6F,
xmp.iid:5C5E8AC43B21DF118E14AB5AA9CB3F6F, xmp.iid:364D8BC67521DF118E14AB5AA9CB3F6F,
xmp.iid:374D8BC67521DF118E14AB5AA9CB3F6F, xmp.iid:384D8BC67521DF118E14AB5AA9CB3F6F,
xmp.iid:1385E8E69824DF119D46DE32DF47E67A, xmp.iid:1485E8E69824DF119D46DE32DF47E67A,
xmp.iid:1585E8E69824DF119D46DE32DF47E67A, xmp.iid:1685E8E69824DF119D46DE32DF47E67A,
xmp.iid:846996C39F24DF119D46DE32DF47E67A, xmp.iid:856996C39F24DF119D46DE32DF47E67A,
xmp.iid:866996C39F24DF119D46DE32DF47E67A, xmp.iid:876996C39F24DF119D46DE32DF47E67A,
xmp.iid:D3719246912ADF11AFA6E6294AFB0521, xmp.iid:D4719246912ADF11AFA6E6294AFB0521,
xmp.iid:F1AE3C17572BDF1197D5DCFA11AF9412, xmp.iid:F2AE3C17572BDF1197D5DCFA11AF9412,
xmp.iid:2068DB752C2EDF11A84DC5F88F067B6A, xmp.iid:2168DB752C2EDF11A84DC5F88F067B6A,
xmp.iid:593550E0F730DF11B0B9918CD7D44B50, xmp.iid:5A3550E0F730DF11B0B9918CD7D44B50,
xmp.iid:E59510561A6CDF11A4818CCF7EDFE615, xmp.iid:E69510561A6CDF11A4818CCF7EDFE615,
xmp.iid:12250FFFA871DF118A6AE0FCC3D6E0B3, xmp.iid:13250FFFA871DF118A6AE0FCC3D6E0B3,
xmp.iid:14250FFFA871DF118A6AE0FCC3D6E0B3, xmp.iid:15250FFFA871DF118A6AE0FCC3D6E0B3,
xmp.iid:4601A8BE14ABDF1183A2BE947806EC8B, xmp.iid:4701A8BE14ABDF1183A2BE947806EC8B,
xmp.iid:82ADCB15B403E0118920DC76910A31A0, xmp.iid:83ADCB15B403E0118920DC76910A31A0,
xmp.iid:84ADCB15B403E0118920DC76910A31A0, xmp.iid:B0F060712F09E011A376BF5163C1EA45,
xmp.iid:B1F060712F09E011A376BF5163C1EA45, xmp.iid:B2F060712F09E011A376BF5163C1EA45,
xmp.iid:B3F060712F09E011A376BF5163C1EA45, xmp.iid:A3975BC55C43E0119CDDC27F8C0DC1DB,
xmp.iid:A4975BC55C43E0119CDDC27F8C0DC1DB, xmp.iid:33A96DDD4444E011BA44C4547900889D,
xmp.iid:34A96DDD4444E011BA44C4547900889D, xmp.iid:35A96DDD4444E011BA44C4547900889D,
xmp.iid:36A96DDD4444E011BA44C4547900889D, xmp.iid:37A96DDD4444E011BA44C4547900889D,
xmp.iid:38A96DDD4444E011BA44C4547900889D, xmp.iid:39A96DDD4444E011BA44C4547900889D,
xmp.iid:3AA96DDD4444E011BA44C4547900889D, xmp.iid:3BA96DDD4444E011BA44C4547900889D,
xmp.iid:3CA96DDD4444E011BA44C4547900889D, xmp.iid:76B300A74E44E011BA44C4547900889D,
xmp.iid:77B300A74E44E011BA44C4547900889D, xmp.iid:FD6F79CF1648E011BFA6AC29DFBD2A96

CreatorTool = Adobe InDesign CS4 (6.0.6)



Postscriptum.

Re: TIA 2011 notification for paper 1

Friday, September 9, 2011 6:40 AM

From: "Dr. Andrej Poleev" <andrepoleev@yahoo.com>

To:

tia2011@easychair.org, msnyder@tiaonline.org, cs.tia2011@inalco.fr, kyo@p.u-tokyo.ac.jp,
Pierre.Zweigenbaum@limsi.fr, lupe@fi.upm.es, alcina@trad.uji.es,
sophia.ananiadou@manchester.ac.uk, aussenac@irit.fr, caroline.barriere@crim.ca,
rita.temmerman@ehb.be, s.sharoff@leeds.ac.uk, secretariat_presidence@inalco.fr,
mvalette@inalco.fr, farid.cerbah@dassault-aviation.fr

Dear idiots within TIA 2011 programme committee,

you are too late with your foolish and useless evaluation of my manuscript that is already published by Journal Enzymes ISSN 1867-3317 and in Scientific and Technical Information Processing, 2011, Vol. 38, No. 2, pp. 119–122. © Allerton Press, Inc., 2011, <http://www.springerlink.com/content/p7567mq32v4h0713/>

Fuck you!

Dr. Andrej Poleev
redactor

--- On Thu, 9/8/11, TIA 2011 <tia2011@easychair.org> wrote:

Dear authors,

The TIA 2011 programme committee would like to thank you for your submission to our conference. We regret to inform you that your paper was not selected for presentation in TIA 2011 conference.

Due to a number of submissions of high quality and merit, our decision was exceedingly hard to make. Unfortunately, the conference limitation only allow for a limited number of papers to be presented in comparison to the submissions we received. Each submission was reviewed by multiple reviewers, and the programming committee carefully examined the review comments. As you can imagine, it was a difficult task to make our final selections, as each submission has its own merits and importance.

We would like to express our appreciation of your interest in TIA 2011.

Yours sincerely,

TIA 2011 Programme Committee

<http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=15439©ownerid=370>

----- REVIEW 1 -----

PAPER: 1

TITLE: Universal Metadata Standard

AUTHORS: Andrej Poleev

OVERALL RATING: -2 (Reject: I believe this paper should not be on the program)

Appropriateness for the conference: 2 (Probably not appropriate)

Clarity: 2 (Sometimes unclear)

Meaningful Comparison to State of the Art: 2 (Only partial awareness and understanding of related work)

Originality: 1 (Significant portions have actually been done before or done better)

Soundness/Correctness: 2 (Troublesome: There are some ideas worth salvaging here, but the work should really have been done or evaluated differently)

Impact of ideas, results or software: 2 (Marginally interesting: May or may not be cited)

SHORT SUMMARY

The paper proposes a novel semantic way of associating documents by similarity with a neuronal network (based on association of notions). Pointing out the incoherence of existing descriptive systems and languages, and criticizing the biases of ontologies, the author claims that a new « Universal Metadata Standard » is urgently needed.

GENERAL COMMENTS AND JUSTIFICATION OF SCORES

The paper intention deserves interest, but the arguments are naive and the general view of the problems doesn't face the complexity of the semantic issue (for example, no difference between metalanguage and natural language). The utopy is a bit outdated. There is a huge gap between the question raised and the proposition.

----- REVIEW 2 -----

PAPER: 1

TITLE: Universal Metadata Standard

AUTHORS: Andrej Poleev

OVERALL RATING: -3 (Strong reject: I'll fight for this paper not to be on the program)

Appropriateness for the conference: 1 (Certainly not appropriate)

Clarity: 1 (Much of the paper is confusing)

Meaningful Comparison to State of the Art: 1 (Little awareness of related work)

Originality: 2 (Pedestrian: Obvious, or a minor improvement on familiar techniques)

Soundness/Correctness: 2 (Troublesome: There are some ideas worth salvaging here, but the work should really have been done or evaluated differently)

Impact of ideas, results or software: 1 (Will have no impact on the field)

The paper is in Russian, and it does not offer any reasonable contribution to research in this field.

TIA 2011 : 9th International Conference on Terminology and Artificial Intelligence

Link: <http://tia2011.crim.fr/Calendar.html>

When

Nov 8, 2011 - Nov 10, 2011

Where

Paris, France

Submission Deadline

Jul 10, 2011

Notification Due

Sep 8, 2011

Final Version Due

Sep 26, 2011

Categories computational linguistics machine translation artificial intelligence ontology

Call For Papers

=====

9th INTERNATIONAL CONFERENCE ON TERMINOLOGY AND ARTIFICIAL INTELLIGENCE

9ème conférence internationale Terminologie et Intelligence Artificielle

Multilingual terminology at the crossroads

Paris, 7-9 November 2011

EXTENDED DEADLINE FOR PAPERS: 10 July 2011

<http://tia2011.crim.fr/>

Submission: <https://www.easychair.org/conferences/?conf=tia2011>

=====

ABOUT TIA 2011

It is widely held that we are living in an era in which global communication technology has enabled us to transcend borders and that we are witnessing the advent of a global community. Although this oft-repeated claim is not without foundation, it is also an undeniable reality that one of the greatest obstacles to global communication -- the language barrier -- remains almost intact and manifests itself in every facet of global communication.

Parallel to this, certain trends - which perhaps emerged quite a while ago - have intensified to such an extent that they have become

issues of serious concern. One is the fragmentation not only of scientific subjects but also of non-scientific activities, and the accompanying subdivision of communications, even within a single language. For example, mathematicians working on different topics use different terminology sets, and the intersection between the jargon used by gamers and that used by musicians is minimal; even among gamers, different jargons are used depending on the type and genre of the game. The other is the massively reduced life expectancy of terms in cutting-edge spheres. It is not uncommon to find that a set of terms one learned several years ago is no longer in use.

While multilingual terminology is a well-established research topic, these changes in the environment in which the topic is situated have given it a new dimension and increased the challenges involved in researching it. It is against this backdrop that we chose multilingual aspects of terminology as the central topic for TIA 2011, while at the same time welcoming a broad range of related topics.

TIA 2011 will be the ninth event in the successful series which started out as a predominantly francophone conference. The conference has gradually attracted a wider audience and over the past six years it has grown to be a truly international conference. Welcome to TIA 2011!

TOPICS

Papers can address theoretical questions or methodological aspects. Interdisciplinary work which emphasizes the convergence and co-operation of multiple disciplines dealing with terminological issues is welcome.

Papers addressing specific issues linked to the topic of multilingual terminology will be particularly welcome, but we are also soliciting original contributions on any of the following topics with a theoretical, methodological or practical orientation:

Terminology acquisition and management

- * term acquisition
- * terminology structuring: acquisition of relations
- * terms as named entities
- * terminology evolution
- * terminology evaluation
- * multimodal information management in terminology

Terminology and knowledge representation

- * ontology construction from terminological resources
- * status of linguistic semantics and ontology
- * knowledge evolution and terminological dynamics

Terminology for applications

- * terminology in industrial applications
- * terminology in education
- * terminology for human translation
- * terminology for machine translation
- * terminology for natural language processing

IMPORTANT DATES

10 July 2011 Deadline for submission

8 September 2011 Notification

26 September 2011 Final camera-ready paper due

7-9 November 2011 Conference

FORMAT OF SUBMISSIONS

TIA 2011 solicits both regular papers, which present significant work, and short papers, which typically present work in progress or a smaller, focused contribution. The program committee may decide to change the category of a paper before acceptance.

Regular papers, in English or French, should not exceed 7 pages in ACL-HLT 2011 format (see below for LaTeX and MS Word style files), including figures, examples and references. The mode of presentation of long papers (orally or as posters) will be decided based on the nature rather than on the quality of the work.

Short papers should not exceed 4 pages in the same format as regular papers. They will also be presented orally or as posters at the conference depending on the nature of the work.

See detailed authors instructions at:

<http://tia2011.crim.fr/Submission.html>

Papers must be submitted electronically in PDF format at

<https://www.easychair.org/conferences/?conf=tia2011> .

PROGRAM COMMITTEE

Presidents

Kyo Kageura (University of Tokyo, Tokyo, Japan)

Pierre Zweigenbaum (LIMSI-CNRS & CRIM-INALCO, Paris, France)

Contact: cs.tia2011@inalco.fr

Members

Guadalupe Aguado de Cea (Universidad Politécnica de Madrid, Spain)

Amparo Alcina (Universitat Jaume-I, Castellón de la Plana, Spain)

Sofia Ananiadou (NaCTeM, Manchester, UK)

Nathalie Aussenac-Gilles (IRIT, Toulouse, France)

Caroline Barrière (CRIM, Montréal, Canada)

Maria Teresa Cabré (Universitat Pompeu Fabra, Barcelona, Spain)

Farid Cerbah (Dassault Aviation, Paris, France)

Jean Charlet (AP-HP & INSERM, Paris, France)

Anne Condamines (CLLE-ERSS, Toulouse, France)

Lyne Da Sylva (EBSI, Montréal, Canada)

Béatrice Daille (LINA, Université de Nantes, Nantes, France)

Valérie Delavigne (Institut national du cancer, France)

Pascaline Dury (Université Lyon 2, Lyon, France)

Fidelia Ibekwe-San Juan (Université Lyon 3, Lyon, France)

Kyo Kageura (University of Tokyo, Tokyo, Japan)

Olivia Kwong (City University Hong Kong, Hong Kong, China)

Marie-Claude L'Homme (OLST, Université de Montréal, Canada)

Adeline Nazarenko (LIPN, Université Paris 13, Villetaneuse, France)

Mihako O'Hagan (Dublin City University, Dublin, Ireland)

Pascale Sébillot (IRISA, Rennes, France)

Serge Sharoff (University of Leeds, Leeds, UK)

Monique Slodzian (ERTIM-INALCO, Paris, France)

Sylvie Szulman (LIPN, Université Paris 13, Villetaneuse, France)

Koichi Takeuchi (Okayama University, Okayama, Japan)

Rita Temmerman (Erasmushogeschool, Bruxelles, Belgium)

Yannick Toussaint (LORIA, Nancy, France)

Spela Vintar (University of Ljubljana, Ljubljana, Slovenia)

Pierre Zweigenbaum (LIMSI-CNRS & CRIM-INALCO, Paris, France)

ORGANISING COMMITTEE

TIA 2011 is organised by the Multilingual Engineering Research Centre
CRIM/ERTIM (EA2520) of INALCO.

Presidents

Mathieu Valette (ERTIM-INALCO, Paris, France)

Monique Slodzian (ERTIM-INALCO, Paris, France)

Related CFPs

IJRA 2011 ACTA Press - International Journal of Robotics and Automation

ECAI 2012 20th European Conference on Artificial Intelligence

PAKDD 2012 The 16th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)

FLAIRS 2012 The 25th Florida Artificial Intelligence Research Society Conference

PAKDD 2012 The 16th Pacific-Asia Conference on Knowledge Discovery and Data Mining

Artificial Intelligence: Turing 2012 Call for Book Chapters: Artificial Intelligence, Evolutionary Computation and Metaheuristics (AIECM) --In the footsteps of Alan Turing (Turing 2012)