

Анализ на семантични технологии за математически текстове

Силвия Върбанова, Юлиана Дошкова-Тодорова

Abstract: *This article describes the capabilities some Semantic Web technologies that are not yet popular and still not yet widely used. Semantic Web technologies are revolutionizing the Web and actively participate in the exchange of information. "Semantic Technologies" is a general term for any software that includes some form and degree of understanding of the meaning of information.*

Key words: *Semantic Web, Mathematical Markup Language, Content-MathML, OpenMath, mathematical formulas in Web, semantic search engine*

ВЪВЕДЕНИЕ

В концепцията за Семантичен Уеб се предлага към обектите в Интернет да се добави формална структура и описание на значението им, наречено семантично описание. При това семантиката и връзките между обектите да са отделени от данните. Семантичната информация е насочена най-вече към машините и организира по-добре работата им по интерпретацията, преработката и обмена на данни. Семантичният Уеб е проект, който има за цел да представя данните в уеб страниците по такъв начин, че те да са разбираеми едновременно за хора и за компютри [3]. Математическите публикации също би следвало да отговарят на тези условия и това се постига чрез специализираните математически технологии от Семантичния Уеб.

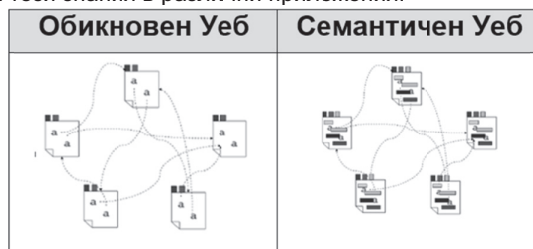
ИЗЛОЖЕНИЕ

1. Семантични данни

Инициативата за Семантичен Уеб стартира през 1998 г. Оттогава се разработват нови езици, актуализират се стандарти, създават се семантично описани бази данни и архиви и техният брой непрекъснато се увеличава. Семантичният Уеб е среда, в която човек и машина могат да комуникират на семантична основа.

Семантичният Уеб представлява мрежа от информационни възли – както и Обикновеният Уеб, но тук те са свързани помежду си по такъв начин, че съществуващата информация може да се обработва от компютри (фиг. 1). Обикновените хипервръзки нямат семантика и не описват взаимовръзките между обектите, а само насочват към допълнителна информация.

Семантичните описания са свързани с понятията „семантични данни“ и „структурирани данни“ и в момента се използват в системи за управление на БД и различни приложения. Към семантичните описания спадат и метаданните – информация за специфичното съдържание на данните. Реализацията на Семантичния Уеб се свързва обаче не само с натрупването на метаданни, а и с машинното им интерпретиране, извеждането на логически връзки и изводи и интегрирането на тези знания в различни приложения.



Фигура 1. Обикновен Уеб и Семантичен Уеб

„Семантични технологии“ е общ термин за всеки софтуер, който притежава някакъв вид и ниво на разбиране на смисъла на информацията. Семантичните технологии могат да революционизират вземането на решения и да участват активно в обмена на информация.

Архитектурата на Семантичния Уеб включва няколко различни слоя [10], всеки от които се реализира с различни технологии. Основните сред тях са eXtensible Markup Language (XML), Resource Description Framework (RDF) и Web Ontology Language (OWL).

Синтаксисът на XML позволява избор на собствени тагове, с които се описват данни. Така значението на XML-документите е интуитивно ясно и таговете, които могат да са думи от естествен език, задават семантиката. След това програмите могат да използват тези тагове по различни начини. XML позволява на потребителите също да добавят произволна структура към документите, без тя да определя значението на всички структури.

RDF е език за представяне на информацията, добавяща семантика в уеб документ. Тя позволява кодирането и обмяната на структурирани метаданни и е машинно-разбираема. Търсещите машини, информационните агенти и браузърите могат да използват тази семантична информация. RDF е независима имплементация, написана по правилата на XML-синтаксиса.

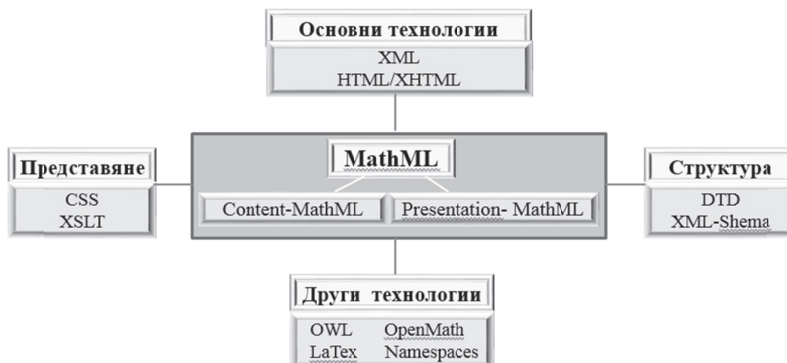
Съществуват още технологии за разработване на Семантичен Уеб като Dublin Core и RDF Schema, а други все още са в процес на разработка.

Част от Семантичния Уеб е изграждането на семантични мрежи. Те дават информационния модел на предметната област, която представляват, и включват свързаните понятия и отношенията между тях. Семантичните мрежи са съвременен подход за представяне на знания.

2. Семантични технологии за математически текст

Математическите формули, както всички обекти в Семантичния Уеб, би трябвало да се възприемат по два начина. От една страна те са поредица от символи и трябва правилно да се изобразят, например в браузър. От друга страна би трябвало да се даде възможност за вникване в техния смисъл.

Най-новите семантични уеб технологии за математически текстове в Уеб целят интегрирането на математически символи и формули, но и позволяват добавяне на семантика към структурните им описания. На Фигура 2 са представени обикновени и семантични технологии и връзките между тях.



Фигура 2. Математически технологии от Семантичния Уеб

Mathematical Markup Language (MathML) е XML-базиран език за описание на математически изрази [6]. Позволява включване на математически символи в уеб документ, добавяне на логическа структура към тях и описание на смисъла на елементите в математическите формули. Първоначалната версия е представена през 1999 г., а през 2014 г. е разработена трета версия. MathML има две части, всяка с различно предназначение и различно множество от маркиращи тагове, но с възможност за съвместно използване. Те са Presentation-MathML и Content-MathML. Presentation-MathML е ориентиран към графичното представяне на данните, т.е. визуализация, позициониране и последователност на математическите символи. Content-MathML описва семантиката и структурата на математическите формули, например функции, оператори, променливи, константи. Така се определя по-добре значението на формулата и тя може да се използва в различни програми [2] и сайтове за обучение [1]. MathML-код може да се създава чрез обикновен текстов редактор или чрез специализирани инструменти [9], които значително улесняват разработчиците.

OpenMath е стандарт [8] за описание на съдържанието на математически формули, който е фокусиран върху кодиране на техния смисъл. Основан е на езика XML. Описанието на логическата структура на формулите се извършва със специален тип данни, наречени OpenMath-обекти, а задаването на семантиката се осъществява чрез колекции от математически понятия, наречени OpenMath-речници. Осигурено е съвместното използване на OpenMath и езика MathML.

OMDoc е създаден за работа с по-големи, в сравнение с обикновените формули, математически структури – дефиниции, теореми, доказателства, теории. При това за описанието на формулите могат да се използват стандартите Content-MathML или OpenMath, а за представянето им се налага конвертиране в Presentation-MathML.

Според Тим Бърнърс-Лий [3] семантичните уеб технологии, свързани с математически текстове, ще направят Мрежата много по-добра по отношение на образователни, научни и технически материали.

3. Семантични математически търсещи машини

При семантичното търсене се акцентира върху значението на ключовите думи, отчитайки свързани по смисъл знания, и затова получените резултати са по-коректни. Семантичните търсещи машини са базирани на семантичните мрежи и онтолозиите и използват специално създадени алгоритми.

Търсенето на математически формули не е стандартен проблем [5]. Трябва да бъде осигурено намирането на дадена формула в друга формула и да се съобрази, че математическите нотации са контекстно-зависими и различни математически записи и означения могат да имат еднакъв смисъл.

Сред най-популярните математически семантични търсачки са MathWebSearch (MWS), EgoMath, LeActiveMath, Symbolab и Wolfram Alpha.

MWS [7] е семантична търсачка на формули, която се базира на съдържанието им и индексването на математически понятия. Потребителите могат да търсят формули, написани на Content-MathML или на друг специализиран език, който позволява конвертиране към Content-MathML. Добавен е и визуален редактор, с който може директно да се създаде формула, и тя след това да се търси.

В други проекти за семантични търсещи машини като EgoMath и LeActiveMath се извършва търсене не само сред формули, но и в текстови документи, съдържащи математически формули.

Семантичните търсачки могат да се използват от математици, информатици и други учени, както и в обучението по математика на студенти и ученици. Те могат да помогнат в търсенето на математически публикации, съдържащи дадена формула или близка на нея – близка теоретично и семантично, а не толкова близка визуално.

ЗАКЛЮЧЕНИЕ

Семантичните езици и технологии дават частична формализация на съдържанието на уеб документите и така осигуряват управлението на данните от семантичните уеб приложения. Семантичният Уеб предполага тълкуването на данни и метаданни и това винаги се извършва в определена област и по отношение на даден модел и в известен контекст. Семантичните анотации в математическите текстове осигуряват достъп до информацията и нейното значение и могат да се съчетават с онтолозиите и базите от знания.

Семантичният Уеб е сред новите разработки на организацията за уеб стандарти W3C. Широкото разпространение на Семантичния Уеб би довело до „революция“ в Интернет според Тим Бърнърс-Лий. Същевременно семантичните технологии са по-сложни за реализация и поддръжка и е нужно повече време за тяхното утвърждаване. Актуални и особено важни са и въпросите за сигурността [4] и коректността на данните и метаданните.

ЛИТЕРАТУРА

[1] Върбанова, С., Ю. Дошкова-Тодорова, MathML в информационната система за комбинаторни обекти DVCobj, Сборник научни трудове на Русенския Университет, Русе, 2014.

[2] Дошкова-Тодорова, Ю., Новите технологии за математически уеб сайтове, Сборник доклади „Иновации в програмните технологии, алгоритми и обучението във висшите училища, свързано с тях“, Велико Търново, 2010.

[3] Berners-Lee, T., M. Fischetti, Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor, Harper San Francisco, 1999.

[4] Bogdanova, G., T. Todorov, N. Noev, Protection of Semantic Organized Data Encryption of RDF Graph, Digital Preservation and Presentation of Cultural and Scientific Heritage (DiPP14), Veliko Tarnovo, Bulgaria, 18-21 Sept, 2014, Vol. 4, No. 1, pp.183-189, ISSN: 1314-4006, 2014.

[5] Kohlhase, M., I. A. Sucas, A Search Engine for Mathematical Formulae, Proceedings of Artificial Intelligence and Symbolic Computation, Springer Verlag 2006.

[6] Mathematical Markup Language (MathML) – W3C Recommendation, <http://www.w3.org/TR/MathML3/>

[7] MathWebSearch: Searching Math on the Web, <http://search.mathweb.org/>

[8] The OpenMath website, <http://www.openmath.org/>

[9] Tools – Math Working Group, <http://www.w3.org/Math/wiki/Tools>

[10] W3C Semantic Web Activity, www.w3.org/2001/sw

За контакти:

Д-р Силвия Ангелова Върбанова, Катедра “Информационни технологии”, Факултет “Математика и информатика”, Великотърновски университет “Св. св. Кирил и Методий”, e-mail: hotsilverbg@yahoo.com,

Доц. д-р Юлиана Дошкова-Тодорова, Катедра “Информационни технологии”, Факултет “Математика и информатика”, Великотърновски университет “Св. св. Кирил и Методий”, e-mail: doshkova@uni-vt.bg