

# Электронная система на основе MS WSS для гармонизации терминологии аналитической химии\*

© В.И. Широкова, В.П. Колотов, М.В. Аленина

Институт геохимии и аналитической химии им. В.И.Вернадского РАН,  
119991, ГСП-1, ул. Косыгина, 19  
shirokova@geokhi.ru

## Аннотация

Для сообщества российских химиков-аналитиков создана электронная система на основе MS WSS (Windows SharePoint Services), объединяющая географически удаленных пользователей, для совместной работы. Разработанная система интерактивных сайтов предназначена для обеспечения задач как постоянных подразделений Научного совета по аналитической химии (комиссий, отделений), так и временных рабочих групп, создаваемых для решения различных общих задач (например, таких как создание глоссариев, предметных баз данных, организации конференций и др.). Начата работа по созданию онтологии аналитической химии. В качестве первого шага поставлена задача разработки электронного глоссария терминологии. Имея в виду множество существующих источников терминов (ГОСТы, рекомендации Международного союза по теоретической и прикладной химии, системы и др.) и существующие разночтения в их трактовке, ведется работа по созданию всеобъемлющей базы данных терминов по аналитической химии из разных источников для их сопоставления, выявление противоречий или разночтений. Очевидно, что обойтись без публичного обсуждения обнаруженных разночтений и выработки консенсуса профессиональным сообществом химиков-аналитиков невозможно. Функциональность развернутых WSS сайтов попадает точно в цель поставленной задачи. С целью систематизации имеющихся рекомендаций проведен критический анализ структур дескрипторов XML-данных, разработанных различными международными организациями (IUPAC, ASTM и др.) для представления химической и аналитической информации.

## Введение

Опыт эксплуатации российского интернет-портала по аналитической химии [1], показал определенные ограничения, свойственные традиционным сайтам, администрируемым централизованно. Так например, НСАХ РАН состоит из многочисленных комиссий по направлениям аналитической химии, научных ассоциаций и семинаров, региональных бюро и др. подразделений, имеющих свои задачи. Их эффективная работа подразумевает наличие способов прямой публикации/распространения информации в Интернете, обмена сведениями по актуальным проблемам и др. Таким образом, в профессиональной среде химиков-аналитиков назрела необходимость рациональной интеграции центрального портала с большим количеством сайтов, обеспечивающих децентрализованную поставку и интерактивное управление узкоспециальной информацией с рабочего места пользователя. Для этого была развернута электронная интернет-среда, основанная на использовании службы MS WSS v.2 (Server 2003). Задачей электронной системы является поддержка работы Научного Совета РАН по аналитической химии (НСАХ РАН) и организация эффективной совместной работы членов совета в рамках различных проектов по аналитической химии (подготовка книг, предметных баз данных, проведение конференций, глоссария по аналитической химии для обеспечения «семантического» Интернета, выполнение планов комиссиями и отделениями совета и др.).

## Развернутые интерактивные WSS-сайты НСАХ РАН

Понятие «интерактивный интернет-сайт» подразумевает, что пользователи такого сайта, в соответствии с наделенными правами, могут самостоятельно размещать на них разнообразную информацию (новости, объявления, документы, организовывать и поддерживать форумы и др.) пользуясь привычным интернет браузером. Вместе с тем, вся размещаемая информация на таких сайтах и центральном портале интегрирована, например, для того, чтобы обеспечить работу поисковой системы или обеспечить ав-

---

Труды 10-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2008, Дубна, Россия, 2008.

томатизированную поставку новостей на центральный портал. Задачи поддержания целостности такого централизованного хранилища информации возлагаются на разработанное программное обеспечение.

MS WSS (v.2) отличается гибкостью настройки, многоязычным интерфейсом, тесной интеграцией с MS SQL-сервером, широкой поддержкой возможностей языка XML. Сайты на основе WSS используют развитую технологию активных серверных страниц (ASP.NET) и допускают удаленное управление и программирование с использованием XML-протокола для обмена информацией в распределенных системах. Последняя версия службы (v.3) отличается заметно более развитыми особенностями (наличие master-страниц, поддержкой ведения сетевых дневников – блогов и др.).

Для выполнения этой задачи использован двухпроцессорный сервер (Xeon 1.7 МГц, отказоустойчивый дисковый массив RAID 1 на основе SCSI дисков), операционная система Windows Server 2003 (Standard Edition), включая Интернет-сервер (IIS v.6.0) и сервер баз данных (MS SQL Server 2000 Standard Edition SP3). Установлена служба WSS v.2 (SP2) в конфигурации Server Farm, обеспечивающая масштабируемость приложения. Так как новый сервер вошел в состав уже существующего домена, то конфигурация Server Farm, позволяет использовать один и тот же контент (базу данных) для нескольких виртуальных серверов. Такой сценарий может быть применен, например, для параллельной работы в интернет/интранет сети. После установки WSS, на IIS был создан виртуальный сервер верхнего уровня Научного Совета по аналитической химии РАН, который был расширен для использования WSS. Этому виртуальному серверу была назначена определенная база данных.

Реализованная архитектура системы подразумевает, что созданный корневой сайт является родительским для всех остальных сайтов системы. Далее с использованием администрирования корневого сайта созданы дочерние сайты для всех подразделений ИСХ РАН и временных групп, а также «входной» wss-сайт, который предоставляет список всех дочерних сайтов, включая соответствующие ссылки [2]. Развернутая система включает более 40 WSS-сайтов различного назначения.

Для автоматизации развертывания типовых сайтов с предопределенной функциональностью и структурой были разработаны различные шаблоны для сайтов (например, комиссий, отделений, конференций, работы над книгами и глоссариями). С использованием шаблонов были созданы сайты для всех подразделений совета (см. входной wss-сайт совета [2]). Генерация одного сайта по шаблону занимает менее минуты, а сам сайт обладает достаточно продвинутой функциональностью: размещение документов с сортировкой их по различным темам, размещение новостей, объявлений, интернет-линков, организация тематических интернет-форумов, поддержка списка контактов группы, ор-

ганизация совместной работы (например, задания, контроль их исполнения) и др.

Очевидно, что интерактивный сайт обеспечивает большую открытость в плане размещения информации, что автоматически означает и большую уязвимость для несанкционированного использования. Для этого разработана стратегия децентрализованного администрирования системы с целью обеспечения устойчивого и надежного ее функционирования. Корневой WSS сайт администрируется исключительно разработчиками портала, а каждый дочерний сайт имеет двух администраторов: один является администратором сервера, а второй представляет соответствующую удаленную группу. Функционирование этой системы основано на ролевой системе безопасности, когда каждый сайт (независимо от соподчиненности) имеет свою независимую систему учетных записей, назначающую посетителям определенные права. При попытке изменить информацию сайта (базы данных) система запрашивает учетное имя и пароль и далее проверяет возможность выполнения планируемого изменения в рамках назначенной роли. Анонимный доступ разрешает лишь просматривать информацию и документы. Администратор группы назначает роли посетителям дочернего сайта (Web-мастер, автор или читатель), а также, в случае необходимости (и наличия навыков) позволяет доступ к общей настройке представления информации на дочернем сайте. Аутентификация посетителей для доступа к функциям управления сайтом проводится средствами безопасности ОС сервера, для чего в системе учетных записей создано соответствующее организационное подразделение (ou). Для аутентификации администраторов всех уровней был назначен SSL (Secure Socket Layer) протокол, обеспечивающий шифрование и безопасную передачу данных через брандмауэр, что особенно важно для удаленного администрирования WSS. Для обычных пользователей (анонимных или ролевых) установлена базовая аутентификация, при использовании которой никакого шифрования информационных потоков не проводится. Кроме того, для каждого виртуального сервера назначен изолированный пул приложения (отдельный процесс операционной системы), что потенциально обеспечивает большую отказоустойчивость. В автоматическом режиме организовано резервное копирование содержания всех сайтов (синхронно с резервным копированием базы данных).

В соответствии с архитектурой WSS, информация всех сайтов хранится в единой базе данных, содержащей десятки реляционно-связанных таблиц и большое число хранимых процедур, обеспечивающих целостность данных. Таким образом, интеграция информации в пределах виртуального сервера существует как неотъемлемое свойство службы WSS.

Одной из целей проекта является интеграция разнородных данных или, другими словами, разработки хранилища информации (Data Warehouse)

интернет-портала и WSS-сайтов. При разработке концепции хранилища руководствовались следующими принципами: данные должны быть предметно ориентированными, должны быть маркированы во времени, должны быть потенциально интегрируемы (поддержка отношений). Процесс построения хранилища включал ряд этапов: доступ к гетерогенным источникам данных (включая ресурсы Интернета), их преобразование и размещение в хранилище; создание метабазы, описывающей смысл информации, помещенной в хранилище. Практика объединения разнородной информации показала перспективность использования объектно-ориентированного подхода. Для практической реализации упрощенного хранилища данных в контексте проводимой работы, мы опирались на имеющиеся аппаратные и программные возможности. Использовали язык описания данных XML, обеспечивающий межплатформенную независимость. Удобство использования XML состоит также в том, что SQL-сервер обеспечивает удобное взаимодействие с информацией такого рода. Данные интернет-портала, размещенные на другом SQL-сервере, были представлены в виде XML-структур, проиндексированы и предоставлены SQL-серверу, обслуживающему сайты WSS. Для синхронизации связанных данных, размещаемых на разных серверах, использовали триггеры (программы SQL-сервера, вызываемые при наступлении определенных событий, таких как загрузка, обновление или удаление данных). Таким образом, ключевая информация, размещаемая на многочисленных WSS-сайтах, мгновенно интегрируется и отображается на головном портале, использующем обычную ASP.NET архитектуру.

Результаты эксплуатации развернутых электронных систем показали высокую эффективность разработанных решений, положенных в их основу.

### **Гармонизация терминологии по аналитической химии**

Одной из главных задач, на решение которой нацелена развернутая система, является обеспечение работ по различным общим проектам совета, которые подразумевают вовлечение ученых разных специальностей. Актуальным проектом является гармонизация терминологии аналитической химии, в том числе и в связи с активно обсуждаемой концепцией семантического Интернета, подразумевающей размещение информационно-ресурсов, содержащих структурированную и формализованную информацию, «понятную» компьютерам [3]. Концепция предполагает наличие в сети предметных онтологий (по сути словарей дефиниций, понятий и терминов той или иной области знания) и увязывание размещаемых в Интернете материалов с этими онтологиями с помощью содержательных дескрипторов XML (вместо HTML). Онтологии, в том числе и по аналитической химии, должны создаваться сообществом специалистов. В этой связи использование возможностей WSS-сайтов является, пожалуй,

оптимальнейшим способом организации такой работы.

Специфика аналитической химии состоит в том, что многие ее методы возникли на стыке с другими науками, что привело к определенному взаимовлиянию терминов, появлению терминов-синонимов, трактуемых по-разному в различных областях аналитической химии. Это относится как к англоязычной, так и к русскоязычной терминологии.

В рамках международного сообщества под руководством Международного союза по теоретической и прикладной химии (ИЮПАК) ведется работа как по систематизации терминологии в различных областях химии, так и по ее гармонизации (единству интерпретации синонимов). ИЮПАКом издаются рекомендации по терминологии различных химических дисциплин, в том числе и аналитической химии [4]. В 2004 г опубликован «Русско-английский и англо-русский словарь терминов по аналитической химии» [5], обобщивший все предыдущие публикации. С января 2006 г. введен ГОСТ Р 52361-2005 «Контроль объекта аналитический. Термины и определения», содержащий определения основных терминов аналитической химии [6]. Периодически публикуются переработки и дополнения терминологии. Тем не менее, противоречия в терминологии остаются.

Например, до сих пор перевод такого ключевого термина аналитической химии как «ассигасу» не однозначен. Его трактуют и как «точность» результатов анализа (близость к опорному значению) или как «правильность». А «точность» часто приравнивают к термину «precision». В свою очередь, «precision» в ряде рекомендаций переводится как «воспроизводимость», которая также рекомендуется как «reproducibility»! И подобных примеров можно привести множество. Очевидно, что такая ситуация неприемлема. Заметным пробелом в решении этой проблемы является отсутствие единого электронного источника терминологии по аналитической химии, включающего и ее трактовку на русском языке и английском языках. Этот пробел несомненно, ощущается как научным, так и образовательным сообществами.

Разрешение неточностей и противоречий формулировок терминов, устранение различного рода невязок является лишь начальным шагом гармонизации терминологии, к сожалению, обычно и единственным. Упускается важнейшая особенность терминологии как языка, представляющего сущности той или иной науки и взаимосвязи этих сущностей. Как наука развивается от простого к сложному, так и терминология отражает этот путь, определяя более сложные понятия через существующие более простые. Глоссарий есть список терминов и их определений (дефиниций). Роль последних состоит в толковании терминов. Однако, очень часто толкование терминов делается относительно произвольно, не используя непосредственно термины более общего типа, а применяя другие определения, в лучшем случае близкие синонимы. Иерархию (систему)

терминов в той или иной области, конечно, должен определять эксперт-профессионал, а соответствие дефиниций предложенной системе может оценить специалист по информационным технологиям, оценивая присутствие требуемых базовых терминов в тексте определений.

Таким образом, другая сторона гармонизации терминологии заключается в формировании компьютеро-ориентированной иерархии терминов, устанавливая между ними семантически прослеживаемые материнско-дочерние отношения, свободные от противоречий.

Решение этой задачи является заметным шагом к построению онтологии, которая отличается более высокой степенью абстракции, «склеивая» терминологию в систему знаний (используя понятия объектно-ориентированно программирования, когда каждый объект наделяется наследуемыми свойствами и методами функционирования). Результирующая онтология, размещенная в качестве электронного ресурса в сети, является ключевым узлом семантической сети.

### База данных глоссария

Работа по созданию онтологии аналитической химии начата сообществом химиков-аналитиков в рамках НСАХ РАН.

В качестве первого шага решается задача разработки электронного глоссария терминологии. Структура реляционной базы данных электронного глоссария представлена на рис. 1.

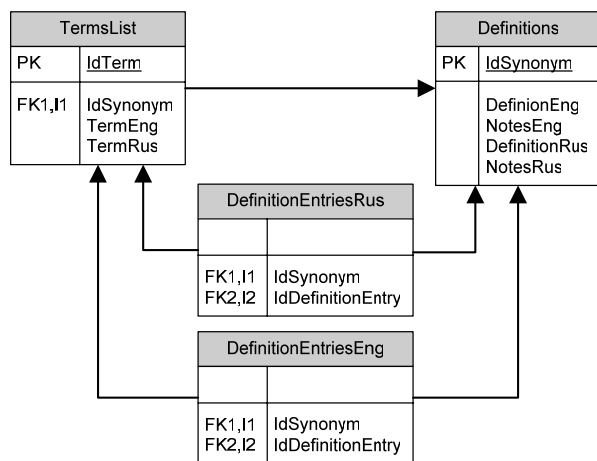


Рис. 1. Структура реляционной базы данных глоссария (PK и FK – первичные и внешние ключи для обеспечения целостности данных и их каскадного изменения/удаления)

Имеется всего две основные таблицы: общий список терминов (TermsList) и список определений терминов (Definitions). Таблица TermsList содержит единственное ключевое поле идентификатора записи (целое число), идентификатор семантической идентичности терминов и собственно термин (на двух языках). Следует отметить важность второго поля. Например, в Международном глоссарии [10]

следующие термины представлены в единой семантической группе с общим определением: «measured quantity value»≡ «measured value of a quantity»≡ «measured value»= quantity value representing a measurement result. В этом случае фрагмент таблицы TermsList выглядит так:

IdTerm	IdSynonym	TermEng	...
...	...	...	...
123	123	measured value	...
124	123	measured value of a quantity	...
125	123	measured quantity value	...

Значение поля IdSynonym будет идентично для всех семантических форм термина и равно идентификатору вхождения первой (обычно наиболее распространенной) формы. Общее определение для всего семантической группы представлено в таблице Definitions, где ключевым полем является идентификатор семантической группы. Такая структура базы данных обеспечивает гибкие возможности для расширения семантических форм и эффективна для оценки семантической идентичности при поиске информации. Две другие таблицы «DefinitionEntries»Eng/Rus» (для англо- и русскоязычной терминологии) предназначены для кодировки вхождения базовых (родительских) терминов в определения более общих (дочерних) терминов. Эти таблицы позволяют построить иерархию и отношения терминов, выявить «висячие» (вне иерархии) термины высокого порядка (заведомо не базовые), определения которых не включают термины более низкой иерархии. Наличие таких случаев может свидетельствовать о неудачно составленном определении термина. Компиляция упомянутых двух таблиц может проводиться автоматически путем последовательного поиска вхождений терминов соответствующего языка в определения терминов (для этой задачи используется хранимая процедура, которая может запускаться триггером при загрузке/изменении термина или его определении). С использованием данных таблиц «DefinitionEntries»Eng/Rus» автоматически строится деревоподобная структура терминологии, которая должна в основном соответствовать иерархии терминов, профессионального сообщества. Гармонизация терминологии достигается при разрешении выявленных противоречий.

### База данных глоссария и WSS-списки

Таблицы на WSS сайтах называются списками. Существуют средства экспорта/импорта таких списков из баз данных (или таблиц Excel). Особенно тесное взаимодействие между списками WSS разработано для MS Access (MS Office2007), где возможен как экспорт/импорт данных, так и динамическая связь между таблицами MS Access и WSS-списками. Однако, обеспечить полноценное пред-

ставление системы реляционно связанных таблиц с помощью списков WSS невозможно (данные всех списков хранятся в единой таблице базы данных WSS). Поэтому вопросы обеспечения целостности данных, включая каскадное обновление/удаление данных, должны решаться особыми средствами.

В разработанной нами системе в качестве основной базы данных используется стандартный MS SQL2000, включая все необходимые средства поддержания целостности (для удобства работы шлюзом к базе данных является проект MS Access). Далее, на сайте WSS созданы два списка, отвечающие основным таблицам глоссария (TermsList и Definitions). Список Definitions включает поле подстановки из TermsList (IdSynonym) для обеспечения ограниченной целостности данных, а TermsList (как любой список) имеет уникальный идентификатор записи.

Как уже упоминалось данные всех списков хранятся в общей таблице WSS, которая называется UserData (или AllUserData в случае WSS v.3), а также ряде вспомогательных таблиц, содержащих идентификаторы списков. С помощью запросов данные из базы данных глоссария экспортируются в базу данных WSS. В результате наполненные списки становятся доступными в Интернете. Специалисты НСАХ РАН могут ознакомиться с системой терминологии и отредактировать для каждой проблемной записи стандартную вложенную форму (файл Excel), где сформулировать проблему и предложения. Непосредственное редактирование полей списка запрещено. На основании предложений администратор модифицирует записи в списке, сопровождая это комментариями в той же форме. По окончании разработки глоссария проводится обновление таблиц исходной базы данных с использованием запроса к WSS-базе данных.

## Заключение

В настоящее время на сайте размещен электронный глоссарий терминов ГОСТ [6]. Ведется работа по обработке терминов ИЮПАК, а также данных из других источников. Определены запросы базы данных и программы по сопоставлению терминов и выявлению противоречий или разночтений с автоматическим размещением на WSS-сайтах для публичного обсуждения, сбора и последующего учета откликов научной общественности и принятия окончательного решения.

## Литература

- [1] Российский Интернет-портал «Аналитическая химия в России»: <http://www.rusanalytchem.org>
- [2] WSS-сайты НСАХ РАН: <http://www.rusanalytchem.org/wss>
- [3] Berners Lee T. Semantic Web Road Map: <http://www.w3.org/DesignIssues/Semantic.html>
- [4] Compendium of Analytical Nomenclature: Definitive rules 1997. Orange Book. 3rd edition Inczedy, J.; Lengyel, T. and Ure, A.M. Blackwell Science, 1998 [ISBN 0-86542-6155]. On-line

version:

[http://www.iupac.org/publications/analytical\\_comp\\_endium](http://www.iupac.org/publications/analytical_comp_endium)

- [5] Русско-английский и англо-русский словарь терминов по аналитической химии / сост. и ред. проф. В.М. Иванов, акад. Ю.А. Золотов. – М. : Изд-во Лаб-пресс, 2004. – 192 с. [ISBN 5-902658-01-2].
- [6] <http://194.67.119.6:83/glossary/>
- [7] What is AnIML?: <http://animl.sourceforge.net/>
- [8] Standard XML data dictionaries for chemistry.: <http://www.iupac.org/projects/2002/2002-022-1-024.html>
- [9] About the IUPAC Compendium of Chemical Terminology (Gold Book): <http://goldbook.iupac.org/about.html>
- [10] International Vocabulary of Metrology – Basic and General Concepts and Associated Terms, VIM 3rd edition, JCGM/WG 2, Document N341. 2007.

## Electronic System based on MS WSS for Harmonization of Terminology in Analytical Chemistry

V.I. Shirokova, V.P. Kolotov and M.V. Alenina

An electronic media for Russian analysts has been developed on the base of MS WSS (Windows SharePoint Services). The media joins geographically remote users for making of common projects. The developed system consists of interactive Internet sites intended both for supporting of needs of the permanent divisions of the Scientific Council on Analytical Chemistry (commissions, divisions) and temporary working groups, which are creating for solution of various common tasks (compilation of glossaries, subject data bases, organization of conferences, etc.). One of the most urgent tasks consists in development of ontology in analytical chemistry. This work has been started recently. Development of electronic glossary of terminology (Russian/English) has been selected as the first step. Taking into account that now various official sources of terms are available (Russian standards, IUPAC recommendations, etc.) which often contradict each other, it was decided to create a comprehensive data base covering all sources. On the next step of the work the terms and their definitions should be compared to reveal conflicting variants of reading. The final phase implies elaboration of consensus of the whole professional community of the Russian analysts concerning the terms and definitions. It is clear that just public discussions via common electronic media will allow resolve such great task. It has been proved that functionality of the deployed WSS sites exactly suites the formulated tasks.

For systematization of the recommendations issued by different International organizations (IUPAC, ASTM, others) a critical analysis of the XML-data descriptors for representation of analytical chemistry information has been done too.

---

\* Работа проводится при поддержке Российского фонда фундаментальных исследований (грант № 08-03-00893).