

ИНФОРМАЦИОННЫЕ СИСТЕМЫ: МОДЕЛИ И ТЕХНОЛОГИИ

А. М. Федотов^{1,3}, О. А. Федотова^{2,3}

¹Институт вычислительных технологий СО РАН, 630090, Новосибирск

²Государственная публичная научно-техническая библиотека СО РАН, 630090, Новосибирск

³Новосибирский государственный университет, 630090, Новосибирск

УДК 004.62

Одной из основных современных тенденций социально-экономического развития общества является информатизация всех областей человеческой деятельности. Инструментом реализации этого всеобъемлющего процесса являются **информационные системы** — системы сбора, хранения, обработки, преобразования, передачи и обновления информации. Статья посвящена вопросам создания информационных систем, моделям и архитектуре.

Ключевые слова: модель, информационная система, метаданные, данные, информация, знание.

1 Предназначение информационных систем

В зависимости от конкретной области применения информационные системы (ИС) сильно различаются по своим функциям, архитектуре, реализации. Однако можно выделить одно свойство, которое является общим для всех ИС. Так как любая ИС предназначена для сбора, хранения и обработки информации, то в её основе лежит среда хранения и доступа к данным. Среда должна обеспечивать уровень надежности хранения и эффективность доступа, которые соответствуют области применения ИС.

Информационные системы предназначены для

- организации хранения информации (организация хранилищ, поддержка систем хранения данных);
- управления информацией (добавление, модернизация, изменение данных);
- управления доступом к информации (контроль исполнения правил регламентации доступа к данным), идентификация данных;
- поиска информации;
- извлечения информации и предоставления ее пользователю (компьютерному приложению) в необходимом ему виде (формате);
- визуализации (представление) информации в соответствии требованиями пользователя.

Цель работы ИС — обеспечение конечного пользователя необходимой информацией. Как правило, пользователя не интересует, как устроена технологическая «кухня» информационной системы. Более того, чем меньше эта «кухня» заметна, тем лучше построена та или иная ИС. Опыт показал, что использование проверенных временем и практикой типовых решений позволяет обеспечить эту технологическую «прозрачность» [9].

Кельвин Муэрс в [13] отмечал: *«Проектирование действенной информационной системы требует глубокого понимания психологии пользователей и знания их социального контекста. Мы не можем исходить из предположения, что люди хотят получить от нас информацию, даже когда знаем, что они нуждаются в ней. В основе большинства провалившихся веб-сайтов, невостребованных локальных сетей и никому не нужных интерактивных продуктов лежали ошибочные представления о пользователях и неверные модели их поведения при поиске информации. Пользователи — явление сложное. Пользователи — явление социальное. Информация — тоже».*

Работа выполнена при частичной финансовой поддержке гранта Президента РФ для государственной поддержки ведущей научной школы РФ № НШ-7214.2016.9.

«Таким образом, я предполагаю,» — писал К. Муэрс — «что многие люди, возможно, не захотят получать информацию — и будут избегать пользоваться системой именно потому, что она снабжает их информацией. Обладание информацией причиняет беспокойство и создает неудобства. Всем нам знакомо это чувство. Получив информацию, вы должны прочитать ее, а это не всегда просто. Затем вам нужно постараться ее понять. Понимание может выявить, что вы действовали не в том направлении или что ваша работа была бесполезна. Получается, что отсутствие информации создает меньше проблем, чем ее наличие и использование.»

Востребованность ИС подчиняется двум законам. *Закон первого лица*: «Информационная система любого предприятия работает только в том случае, если ее работа поддержана соответствующими нормативными актами предприятия и с ней работает первое лицо предприятия» и *закон Муэрса* [13]: «Степень использования информации прямо пропорциональна легкости ее получения». Система получения информации окажется невостребованной, если обладание информацией будет вызывать у клиента больше неудобств и беспокойства, чем ее отсутствие [13].

Очевидно, что автоматическая или автоматизированная переработка «информации» возможна лишь при наличии ее описания с помощью некоторого алгоритма, т. е. при наличии формальной модели данных и некоторой системы, которая ее воспринимает. Как отмечал А. А. Ляпунов, «информация всегда относительна, она зависит . . . от того, какой информационной системой она воспринимается» [5].

Эффективная работа ИС возможна только тогда, когда мы располагаем «полноценной информацией»: т. е. «информация» должна быть не только полной, достаточной, но и своевременной, точной и непротиворечивой, а «информация», поступающая через различные по своей физической природе каналы, должна быть и согласованной.

2 Определения информационных систем

В этом вопросе нет единого определения. В обиходе **информационными системами** называют различные комплексы программно-аппаратных средств, позволяющие работать с данными, структурированными при помощи той или иной формальной модели.

При этом до сих пор присутствует путаница с понятиями «информационная система» (ИС) и «архитектура ИС»; она вовсе не безобидна и часто мешает на практике четко определить, что же является предметом разработки в конкретном проекте: ИС, только ее КСА (комплекс средств автоматизации) или автоматизированная система (АС) целиком. Для примера приведем два определения из ГОСТ'ов:

Информационная система: Комплекс, состоящий из процессов, технических и программных средств, устройств и персонала, обладающий возможностью удовлетворять установленным потребностям или целям [1];

Информационная система: система, предназначенная для сбора, передачи, обработки, хранения и выдачи информации потребителям и состоящая из следующих основных компонентов: программное обеспечение, информационное обеспечение, технические средства, обслуживающий персонал [2].

Стоит напомнить, что в 1950-е и 1960-е годы в СССР науки об информации занимали весьма достойное место, хотя их развитие и было затруднено спецификой социального устройства общества. Произошедший в последующие годы разворот к работе с данными и размытие термина «информатика» привели к тому, что многое из накопленного оказалось если не потерянным, то невостребованным, а культура работы с информацией была утеряна. Сегодня для большинства пользователей важнее потребление информационных сервисов, а не обеспечивающие его технологии.

Деструктивная роль информационного шума «многоязычия» осознана еще в сюжете Вавилонского столпотворения. В отличие от социальной среды, где «языки» наряду с информационной несут и социально-психологическую нагрузку, компьютерную среду нужно избавить от «многоязычия». Полисемия (многозначность представления) — враг информационного обмена, фактор внесения искажения и ошибок (семантического шума) на пути передачи информации. Поэтому устранение многозначности является одним из наиболее важных направлений в разработке формальных подходов представления информации и знаний.

Общество уже преодолело первый кризис «многоязычия» информационных технологий в технической сфере: в конце 1990 годов был принят целый ряд стандартов взаимодействия открытых систем OSI, которые сейчас восприняты всеми производителями компьютерного и телекоммуникационного оборудования.

Но наиболее серьезной проблемой является кризис в сфере представления информации. Важной деструктивной особенностью является отсутствие единых общепринятых определений в сфере информационных технологий, когда речь идет об обработке «информации». Прежде всего, потому что со времен А. Н. Колмо-

горова и Клода Шеннона на инженерном уровне произошло смешение понятий, объединение представлений об информации и данных или сигналах, кодирующих эту информацию, и под «информацией» стали понимать, по существу, наборы данных.

До последнего времени, пока ИС были относительно просты, отсутствие четкого разделения на «данные» и «информацию» не имело практического значения. Но с появлением сложных ИС, где функции распределены между человеком и машиной, а также с развитием таких дисциплин, как поддержка принятия решений и управление знаниями, требуются более точные определения базисных понятий: «данные», «информация» и «знание».

На сегодняшний день имеется два определения ИС (технологическое и инженерное):

Информационная система — это набор технологий, направленных на поддержку жизненного цикла «информации» и включающих три основные составляющие процесса: обработку и управление данными, управление информацией и управление знаниями [6];

Информационная система — это программно-аппаратный комплекс, включающий вычислительное и коммуникационное оборудование, программное и лингвистическое обеспечение, информационные ресурсы, а также обслуживающий (системный) персонал.

Часть реального мира, которая моделируется ИС, называется ее *предметной областью*. Поскольку модель предметной области, поддерживаемая информационной системой, материализуется в форме организованных необходимым образом информационных объектов, она называется *информационной моделью* (см. рис. 1). Информационные объекты характеризуются метаданными (данными об объекте), описывающим реальный объект. Информационные объекты могут иметь информационное содержание (контент).

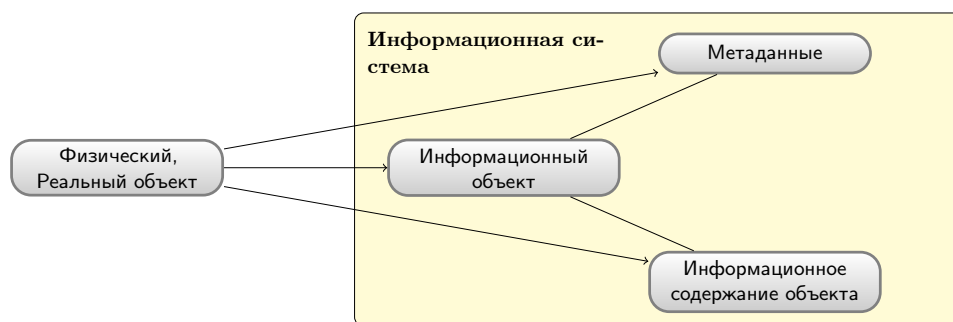


Рис. 1: Информационная модель ИС

Приведенные определения охватывают все классы информационных систем, в частности фактографические системы, которые основаны на технологиях баз данных и оперируют структурированными данными, и документальные системы, оперирующие документами на естественных языках и др.

3 Информационный ресурс

Информационный ресурс — это понятие, включающее любую сущность ИС. В информационном пространстве все сущности (документы, публикации, персоны, события, факты, программы и любые другие сущности реального или виртуального мира) существуют только в форме некоторых информационных объектов. Информационный ресурс — это абстрактное понятие, выражаемое экземплярами одной из своих специализаций. В частности, экземплярами понятия информационного ресурса являются экземпляры информационного объекта любого типа.

Огрубляя сказанное, любая ИС представляет собой систему управления информационными ресурсами с присущими ей функциями (методами), отношениями и связями [9]. Каждый информационный ресурс в соответствии с эталонной моделью (например, OAIS RM [12] или DLRM [10]) (см. рис. 2).

- имеет идентификатор;
- организован в соответствии с описанием ресурса. Ресурс может быть сложным и структурированным, поскольку он, в свою очередь, может состоять из меньших ресурсов и иметь связи с другими ресурсами. С организационной точки зрения, ресурсы могут группироваться в наборы ресурсов, которые рассматриваются как единая сущность;

- может регулироваться функциями, управляющими его жизненным циклом, характеризуется набором присущих ему атрибутов и методов, характеризующих его свойства и связи с другими ресурсами;
- выражается через информационный объект.

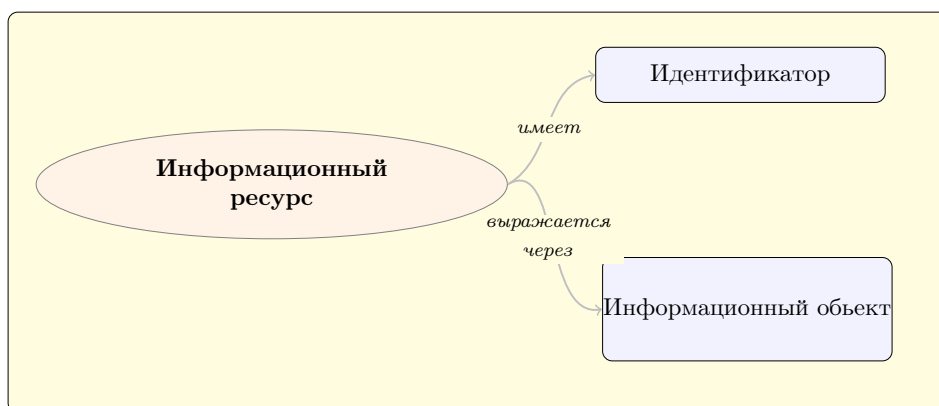


Рис. 2: Определение ресурса в ИС

Реализация информационного ресурса (информационный объект) — это единица информации, представляющая собой уникально именованный набор данных. Каждый информационный ресурс хранится в некотором репозитории как глобально и уникально именуемый набор структурированных данных (сведений о ресурсе, его свойств, атрибутов, методов, связей с другими объектами) и, возможно, содержания, например, одного или более форматов представления ресурса. Эти структурированные данные, описывающие ресурс, а в связи с этим называемые метаданными, используются для получения представления о его свойствах, содержании, структуре, организации поиска, способах использования и т. п.

Репозиторий — средство долговременного и надежного хранения информации, предоставляющее некоторый четко специфицированный способ управления ею (схему данных, модель операций), включающий в той или иной мере способы доступа, выборки и манипулирования информационными ресурсами.

Коллекция — это совокупность информационных объектов (информационных ресурсов), объединенных общими свойствами (например, принадлежностью к одному классу объектов, одинаковой структурой, общей тематической направленностью и т. п.).

В ИС каждому ресурсу соответствует информационный объект, который является традиционным «вторичным информационным объектом», содержащим описание первичного ресурса (каталожная запись), т. е. информационный объект — это объект, который хранит информацию о объектах ИС (физических объектах, ресурсах, информационных объектах). Поскольку информационный объект является ресурсом, то он наследует все вышеперечисленные свойства ресурса, т. е. имеет идентификатор (см. рис. 3).

В свою очередь, каждый информационный объект в ИС состоит из следующих объектов:

- *метаданных* — объект, главная цель которого состоит в том, чтобы дать информацию о ресурсе;
- *аннотаций* — объект, главная цель которого состоит в том, чтобы аннотировать ресурс или его часть. Примеры таких аннотаций включают примечания, структурированные комментарии и связи. Объекты аннотации помогают интерпретировать ресурс, содержат детальные объяснения, либо информацию о том, как можно использовать ресурс [9].
- *информационного содержания* — объект, который может отсутствовать и может использоваться самостоятельно, как первичный информационный объект: например, изображение, полный текст и т. д. (первичный ресурс).

Информационный объект — наиболее общее понятие в системе, представляющий произвольную единицу информации в ИС. Информационные объекты также могут быть сложными объектами и могут быть сгруппированы в коллекции информационных объектов, которые, в свою очередь, тоже являются информационными объектами. Коллекции, в свою очередь, тоже являются информационными объектами, они наследуют все аспекты моделирования информационных объектов и средства их обслуживания, например, они могут аннотироваться.

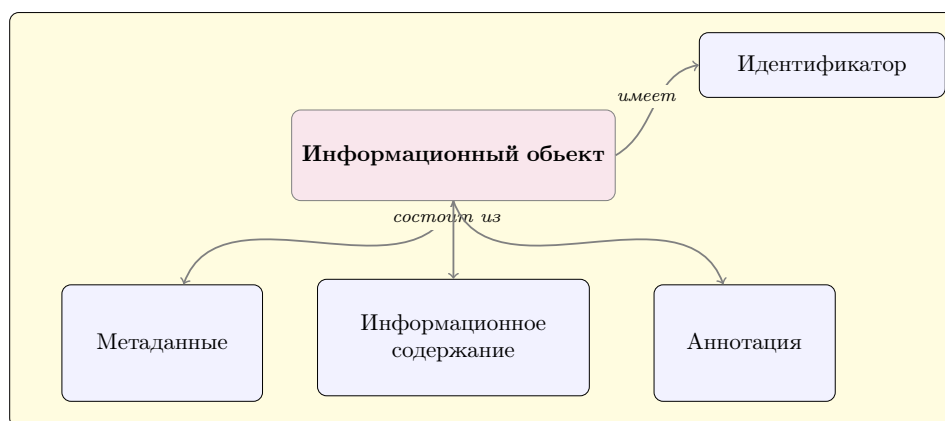


Рис. 3: Модель информационного объекта в ИС

Информационные объекты описывают основные *классы сущностей* научного информационного пространства такие, как публикация, персона, ключевой термин или понятие, словарная статья, факт, функция, организация, пользователь и т. д., а также связи между ними [7, 8].

Одним из фундаментальных понятий в ИС является понятие «документ». Содержание понятия документ многозначно и зависит от того, для каких целей оно используется. Известный бельгийский ученый Поль Отле в начале XX века вводит новое, по сравнению с тем, что было, более широкое понимание понятия «документ». В своем «Трактате о Документации» [14] он определяет это понятие так: *Документ — «материальный объект, содержащий информацию, специально предназначенный для ее передачи в пространстве и времени»*. Важно заметить, что термин «документ» более старый, нежели термин «информация».

В информационной системе *Документ — это целостный информационный объект, представленный в цифровом виде, имеющий некоторый стандартный набор атрибутов и функций и допускающий однозначную идентификацию*.

4 Метаданные

Ключевым моментом в работе с информационными объектами и, в частности, с документами является использование метаданных. Метаданные — структурированная информация, которая описывает, поясняет и указывает местоположение информационного объекта [3]. Метаданные необходимы для решения следующих задач:

- предоставление сведений о документах, для получения представления о их свойствах, содержании, структуре, способах использования и т. д.;
- систематизации информации о документах и ведения каталога системы;
- выбор из множества документов определенного подмножества по формальным признакам и сопоставление документов по формальным признакам;
- внутрисистемные технологические задачи, связанные с обеспечением подготовки документов, размещением документов в информационной среде и т. д.;
- внешние технологические задачи, связанные, прежде всего, с обменом данными с внешними информационными системами.

Метаданные отражают наиболее существенные свойства объекта, имеющие наибольшее значение с точки зрения информационной системы. Метаданные подразделяются на три типа: описательные, структурные и системные.

Описательные метаданные — метаданные которые описывают содержание и свойства информационного ресурса (объекта), например, библиографические данные или аннотация, основная задача которых однозначное представление цифрового объекта для внешнего мира и в различных приложениях.

Структурные метаданные — метаданные которые характеризуют общую структуру информационного ресурса и его компоненты, объем и другие свойства информационного ресурса (объекта);

Системные метаданные (метаданные административные) — служат для обеспечения системы управления информационными ресурсами и администрирования информационных ресурсов, например, даты создания или модификации ресурса, идентификатор владельца и т. п.

Состав имен атрибутов, ограничения, накладываемые на их значения, набор правил, определяющих структурирование атрибутов, их семантику задаются *схемой метаданных*. Правила структурирования метаданных в определенном смысле аналогичны правилам, предлагаемым онтологией для построения отношений между понятиями. Правила, определяющие представление метаданных в информационной системе, а также правила их интерпретации, образуют *формат метаданных*.

Схема метаданных — это набор элементов метаданных, каждый из которых обладает некоторым именем и семантикой, принимает значения с установленной семантикой или значения из *управляемого словаря*, называемым *схемой кодировки*. В соответствии с рекомендациями Dublin Core (DCMI) [11] информационный объект должен обладать базовым набором атрибутов. Набор атрибутов объекта расширяется в зависимости от его класса.

Схема кодировки — система записи или правила анализа значений элементов метаданных. Значение, определенное с помощью схемы кодировки, представляет собой код (символ), выбранный из управляемого (контролируемого) словаря (например, индекс системы классификации или значение из набора предметных рубрик), либо строку определенной структуры (например, "2000-01-01" как стандартное обозначение даты).

Управляемый словарь (Controlled Vocabulary — контролируемый, нормативный, авторитетный словарь) — это список предварительно определенных кодов, терминов, слов, фраз или нотаций, предназначенный для обозначения предметных рубрик или состава допустимых значений атрибутов элементов метаданных. Все коды (термины) в словаре должны иметь однозначное определение.

Особым видом метаданных являются метаданные однозначно характеризующие (идентифицирующие) документы, которые необходимы для систематизации документов и для эффективного поиска, получившие название *авторитетных*.

Особым классом метаданных является метаданные, описывающие отношения и связи между информационными ресурсами — документами.

Отношение — связь между экземпляром некоторой сущности и тем, что с ней соотнесено. По Аристотелю «есть то, что оно есть», лишь «в связи с другим или находясь в каком-то ином отношении к другому». Количество типов отношений в информационной системе определяется, исходя из конкретных целей. В реальном мире их число стремится к бесконечности. С позиций удовлетворения информационных потребностей пользователей нас будут интересовать отношения только между документами, например, «Публикация — Публикация», «Публикация — Персона», «Публикация — Словарная статья», «Публикация — Ключевой термин», «Персона — Словарная статья» и так далее. Связи существуют между всеми классами документов.

В зависимости от условий использования отношения между документами подразделяются на следующие типы: тезаурусные отношения, семантические отношения и ассоциативные отношения:

- *тезаурусные отношения*: отношения, применяемые в описании информационно-поисковых тезаурусов — это иерархические отношения и отношение ассоциации. Основным иерархическим отношением — является родовидовое отношение (родитель-потомок, шире-уже, выше-ниже, часть-целое). Основное назначение установления ассоциативных отношений между документами — указание на дополнительные связи [4]. Тезаурусные отношения специфичны для отношений между ключевыми терминами, значительно реже используются при задании отношений между публикациями и словарными статьями.
- *семантические отношения*: именованные отношения между документами, например, «Персона является автором Публикации»; «Публикация посвящена Персоне»; «Публикация посвящена Факту, описанному в Словарной статье».
- *отношения ассоциации*: отношения между двумя документами которые близки по содержанию, например, ключевые слова в описании Публикации, Персоны, Словарной статьи.

В информационной системе возможно два способа реализации связей (отношений) между документами: жесткие и мягкие. Жесткие связи реализуются средствами СУБД путем ссылок на первичные ключи записи. К сожалению, такой тип связи не защищен от нарушения целостности (в случае неправильного изменения или удаления записи). Мягкие связи реализуются через процедуру поиска соответствий. Такой способ установления связей защищен от любых нарушений целостности БД и достаточно удобен пользователям, поскольку для указания на необходимость связи используются наглядные мнемонические определения.

Заключение

На основе представленной модели ИС в ИВТ СО РАН создана Система Управления Информационными Ресурсами (СУИР), на которой реализованы ряд конкретных ИС. Среди них системы по научному наследию: «Электронная библиотека А.А.Ляпунова», информационно-справочные системы «История информатики в лицах», «История вычислительной техники в лицах» и др.

Список литературы

- [1] ГОСТ Р ИСО/МЭК 12207-99. Информационная технология. Процессы жизненного цикла программных средств. ГОССТАНДАРТ РОССИИ. Москва, 1999.
- [2] ГОСТ 34.003-90. Информационная технология. Комплекс стандартов и руководящих документов на автоматизированные системы. Термины и определения.
- [3] Когаловский М. Р. Метаданные, их свойства, функции, классификация и средства представления // Труды 14-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» — RCDL2012, Переславль-Залесский, Россия, 15-18 октября 2012 г.
- [4] Лукашевич Н. В. Тезаурусы в задачах информационного поиска. М: Издательство МГУ, 2011.
- [5] Ляпунов А. А. О соотношении понятий материя, энергия и информация // В кн.: Ляпунов А. А. Проблемы теоретической и прикладной кибернетики. Новосибирск: Наука, 1980. С. 320–323.
- [6] Советов Б. Я., Цехановский В. В. Информационные технологии: учебник для вузов. М.: Высшая школа, 2005.
- [7] Федотов А. М. Методологии построения распределенных систем // Вычисл. технологии. 2006. Т. 11, Избранные доклады X Российской конференции «Распределенные информационно-вычислительные ресурсы» (DICR-2005), Новосибирск, 6-8 октября 2005 г. С. 3–16.
- [8] Шокин Ю. И., Федотов А. М., Жижимов О. Л., Гуськов А. Е., Столяров С. В. Электронные библиотеки — путь интеграции информационных ресурсов Сибирского отделения РАН // Вестник КазНУ, специальный выпуск. — г. Алматы, Казахстан, Казахский национальный университет им. аль-Фараби. 2005. № 2. С. 115–127.
- [9] Шокин Ю. И., Федотов А. М., Жижимов О. Л., Федотова О. А. Эволюция информационных систем: от Web-сайтов до систем управления информационными ресурсами // Вестник НГУ Серия: Информационные технологии. 2015. Том 13, вып. № 1. С. 117–134.
- [10] D3.2b The Digital Library Reference Model // Funded under the Seventh Framework Programme, ICT Programme — “Cultural Heritage and Technology Enhanced Learning” Project Number: 231551. April 2011. (<http://www.dlorg.eu>) — Full Text: http://db4.sbras.ru/elbib/data/show_page.phtml?22+269.
- [11] DCMI — Dublin Core Metadata Initiative // (<http://www.dublincore.org/>).
- [12] ISO 14721:2012 Reference model for an Open archival information system (RM OAIS) / Recommended Practice: CCSDS 650.0-M-2 (Magenta Book). June 2012. Available at: <http://public.ccsds.org/publications/archive/650x0m2.pdf>
- [13] Mooers C. N. «Mooers» law, or why some retrieval systems are used and other are not // American Documentation. 1960. Vol. 11, No. 3.
- [14] Otlet P. Traité de documentation. Bruxelles: Ed. Mundaneum, 1934.

*Анатолий Михайлович Федотов — чл.-корр. РАН, д.ф.-м.н., гл. науч. сотр. Института
вычислительных технологий СО РАН; Новосибирский государственный университет;
e-mail: fedotov@sbras.ru;*

*Ольга Анатольевна Федотова — науч. сотр. Государственной публичной научно-технической
библиотеки СО РАН; Новосибирский государственный университет;
e-mail: o4f8@mail.ru.*

Дата поступления — 30 мая 2017 г.