

СРЕДСТВА ТРАНСДИСЦИПЛИНАРНОГО ПРЕДСТАВЛЕНИЯ ИНФОРМАЦИОННЫХ РЕСУРСОВ РАЗНЫХ СТИЛЕЙ

Светлана Гайко, Виталий Приходнюк

Аннотация: *Статья касается проблемы эффективного использования разнородных информационных ресурсов (ИР). Огромное количество естественных язычных документов, накопленных в сети, остаются пассивными источниками знаний. Для эффективной работы с такими документами, прежде всего, требуется их структуризация, которая не может производиться вручную, учитывая необходимые для этого трудозатраты. Также необходимо определить способ представления результатов структуризации ИР, т.к. от этого зависят дальнейшие возможности при работе с ними. Важнейшей из таких возможностей является интеграция ИР разных стилей, описывающих различные предметные области (ПДО) в едином информационном пространстве. В статье предлагается модель интеграции разнородных документов, реализующая принцип трансдисциплинарности. Описана технология автоматизированной обработки естественных язычных текстов (ЕЯТ) различных стилей и формирования на их основе интерактивных документов. Рассмотрен механизм отображения их структуры, которым является нарративный дискурс. Представлена математическая модель данной технологии, модель поведения системы (в виде UML-диаграмм), а также архитектура программного средства, созданного на основе описанных моделей.*

Ключевые слова: *трансдисциплинарность, онтология, растущая пирамидальная сеть (РПС), нарративный дискурс, функциональные стили.*

ITHEA классификация: H. Information Systems – H.1 MODELS AND PRINCIPLES – H.1.2 User/Machine Systems.

Введение

В последние десятилетия стала актуальной необходимостью решения большого количества сложных многофакторных проблем в различных сферах человеческой деятельности. В этих условиях повысилась потребность в компьютерных средствах, позволяющих использовать ИР из разных областей знаний. Наряду с этим, ЕЯТ до сих пор остаются самой распространенной формой представления знаний. Такие знания являются пассивными, поскольку большинство документов созданы на основе использования различных информационных технологий и стандартов. В основном, они относятся к классу слабоструктурированных, а по совокупности и характеру изложения, к классу Больших Данных (Big Data) [Mayer-Schönberger, 2013]. Преобразование таких пассивных систем знаний, отображающихся в виде документов, которые содержат описания определенных процессов и их свойств, в интерактивную форму является весьма актуальной проблемой. Однако для этого необходимо реализовать когнитивные процедуры их преобразования, что, как минимум, определяет условия реализации взаимодействия с этими системами, уже активных знаний [Стрижак, Трансдисциплінарна інтеграція, 2014], [Стрижак, 2020].

Таким образом можно выделить основные задачи компьютерной обработки естественнoязычных документов. К ним относятся:

- 1) автоматизированная структуризация разнородных ИР;
- 2) представление результатов обработки в виде интерактивных документов;
- 3) интеграция ИР релевантных задачам конкретного пользователя (эксперта).

Решение поставленных задач связано с рядом технологических и методологических проблем. Это обусловлено как фактом накопления сверхбольших объемов информации, так и высокой интенсивностью процессов сетевого взаимодействия.

Учитывая это, наиболее сложной является именно третья задача интеграции ИР, которая сводится не к простому объединению контекстов нарративов, отвечающих заданной тематике, а обеспечивает активное использование свойств связей между контекстами [Стрижак, Трансдисциплінарна інтеграція, 2014].

В статье описана модель трансдисциплинарного представления разностилевых ИР, лежащая в основе построения информационно-аналитической системы, призванной решать задачи как обработки и структуризации ЕЯТ, так и их конструктивной интеграции.

Обзор литературы

Категория трансдисциплинарности имеет широкую трактовку и применение в современной науке. В сфере информационных технологий данная категория трактуется как гиперсвойство множественной частичной упорядоченности элементов информационной среды [Стрижак, Трансдисциплінарна інтеграція, 2014]. То есть, категория трансдисциплинарности позволяет рассматривать интеграцию информационных ресурсов как некий процесс использования любых релевантных контекстов, и взаимодействия сетевых информационных систем. Данная концепция представлена в работах [Палагин, 2014], [Широков, 2017], [Величко, 2017].

Одним из возможных подходов к реализации трансдисциплинарной интеграции ИР является онтологический подход. Трансдисциплинарные онтологии обеспечивают корректное агрегирование различных тематических процессов путем формирования структурированной

совокупности информационных объектов-концептов ПдО, которые определяются как единый тип данных. Технология их использования в сетевой среде, в которой активируются процессы взаимодействия сложных информационных систем, позволяет установить над ИР, которые активно используются, отношение частичного порядка. Онтологические аспекты трансдисциплинарной интеграции ИР изложены в работе [Стрижак, Онтологические аспекты, 2014]. В работах [Басюк, 2017], [Палагин, 2012] посвященных вопросам онтологического инжиниринга в целом, представляется также широкая библиография по данному вопросу.

Механизмом отображения структуры разностилевых ИР является нарративный дискурс [David, 2012]. Данный механизм подробно описан в статье [Стрижак, 2020]. Суть его заключается в рассмотрении неструктурированных сетевых документов как нарративов, которые представляют собой множество объектов, являющихся концептами, и множество объектов, которые являются классами. Такие объекты, в свою очередь, способны к взаимодействию (свойство дискурса), и это дает возможность отображать связность двух и более нарративов. Построив таксономическую структуру как каждого документа, так и метатаксономию содержания всех используемых документов ПдО, достигается отображение связности документов и процессов взаимодействия с ними.

Входные данные и методы

Модель трансдисциплинарного представления ИР различных стилей является описанием технологии автоматизированной обработки разностилевых документов глобальной среды и представление их в форме, пригодной для дальнейшего превращения в активный формат интерактивных баз знаний. Практическая реализация такой технологии осуществляется на основе WEB-ориентированного сервиса Когнитивная информационная технология (КИТ) «Полиэдр».

Представленная в статье технология предусматривает применение метода построения РПС [Гладун, 2004], а также метода рекурсивной редукции [Приходнюк, 2018].

Для построения онтологического классификатора стилей экспертным образом был обработан массив научных, учебных, законодательных, ведомственных, публицистических и, частично, художественных документов. В результате была построена таксономическая структура функциональных стилей языка, а также определены атрибуты объектов, относящихся к основным стилям.

Основная часть

Модель трансдисциплинарного представления ИР различных стилей состоит из математической модели и модели поведения системы (в виде UML-диаграмм), а также архитектуры программного средства, созданного на ее основе.

Математическая модель

Поскольку механизмом отображения структуры разностилевых ИР является нарративный дискурс, прежде всего, необходимо формализовать данное понятие.

Как отмечалось выше, естественноразличные документы являются пассивными системами знаний. Они, фактически, представляют собой определенные тексты, которые отображаются в виде последовательного изложения определенных концептов, устойчиво заданных условий их существования, описаний их свойств и функциональностей и тому подобное. Их можно представить в виде:

$$O_{nr} = \langle X(K), p \rangle, \quad (1)$$

где O_{nr} – документ с последовательно определенными описаниями;

X – концепты;

K – описания концептов (контексты);

p – отношение строгого порядка, определяет условия существования концептов X в тексте.

То есть, нарративом является текст, который пассивно излагает определенную систему знаний, и имеет признаки онтологической системы, в данном случае, имеет выколотое значение онтологии [Стрижак, Онтологические аспекты, 2014], [Басюк, 2017], [Палагин, 2012], [Палагин, 2016], [Приходнюк, 2018].

Операциональность произвольного текста вида (1) определяется следующими гиперсвойствами [Стрижак, Трансдисциплінарна інтеграція, 2014], [Стрижак, 2020]: рефлексия – \mathfrak{R}_f , рекурсия – \mathfrak{R}_k , и редукция – \mathfrak{R}_d . Когнитивный характер указанных гиперсвойств реализуется в их интерпретации гиперфункциями типа: анализ больших данных, их структурирование, синтез, выбор и принятие решений, которые согласно [Amit, 2005] являются когнитивными.

Данные гиперсвойства образуют определенное замкнутое множество $\mathfrak{R3}$, которое обеспечивает связывание и динамическое изменение упорядочения контекстных описаний нарратива O_{nr} [Malyshevsky, 1998].

Это значит, что выколотую онтологию вида (1) можно расширить множеством $\mathfrak{R3}$ и гиперфункцией ее интерпретации $F_{\mathfrak{R3}}$:

$$O_{nr} = \langle X(K), p \rangle \rightarrow O_{nd} = \langle X(K), \mathfrak{R3}, F_{\mathfrak{R3}} \rangle \quad (2)$$

Через преобразование (2), подстановкой множества гиперсвойств $\mathfrak{R3}$ и функциональным расширением $F_{\mathfrak{R3}}$, вместо отношения строгого упорядочения p можно получить онтологическую систему O_{nd} , которая

обеспечивает различные виды связей между концептами X и контекстами K .

Автоматизированная обработка разностилевых документов состоит из трех этапов:

- классификация ЕЯТ – определение его стиля;
- идентификация информации из текста согласно определенного стиля;
- трансдисциплинарная интеграция информации, содержащейся в различных обработанных текстах.

Для корректной обработки текста согласно его стиля необходимо иметь описания собственно существующих стилей. Естественным способом представления такой классификации является онтологический реестр стилей в виде онтологии O_s (рис. 1).

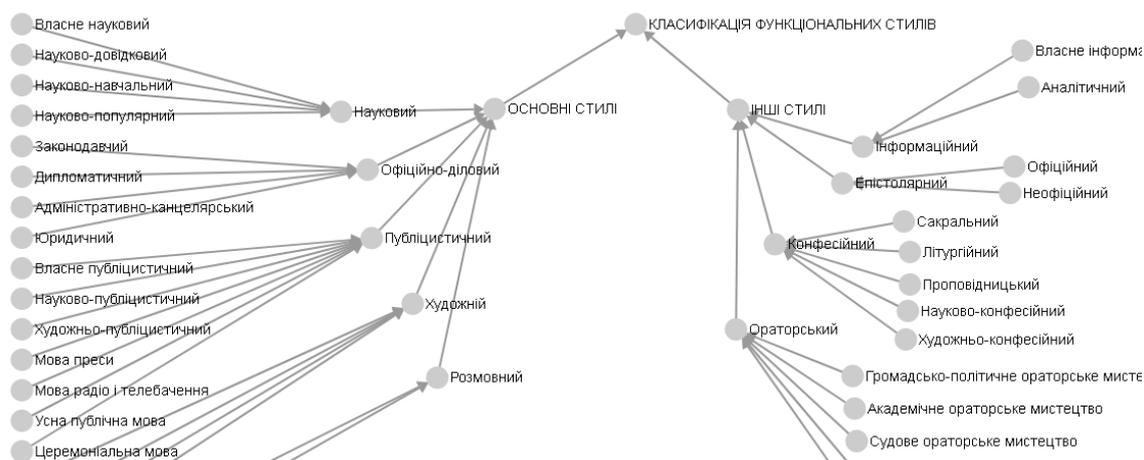


Рисунок 1 Онтологический реестр стилей

Поскольку данный реестр является онтологией, он может быть представлен как:

$$O = \langle X, R, A' \rangle,$$

где X – множество концептов, представленных собственно стилями и группами стилей;

R – множество связей между концептами;

A' – атрибуты концептов, которые задают классификационные характеристики соответствующих стилей.

Задача классификации текстов по стилю решается путем построения РПС. Онтология стилей O_s может быть преобразована в РПС. Для этого нужен промежуточный шаг – выделение таксономии из онтологии:

$$O_s \xrightarrow{G_T} T_s$$

Таксономия T_s имеет структуру:

$$T_s = \langle X_T, R_T \rangle$$

где X_T – множество объектов, принадлежащих таксономии;

R_T – множество связей, содержащихся в таксономии.

Множество X_T состоит из объектов X_s начальной онтологии O_s , а также множества объектов, представляющих классификационные свойства соответствующих стилей (построенных на основе атрибутов A'_s объектов X_s):

$$X_T = X_s \cup X(A'_s)$$

Множество R_T содержит взаимосвязи между объектами X_T , представляющие стили, и объектами $X(A'_s)$, которые представляют классификационные характеристики объектов:

$$R_T = R(X_s, X(A'_s))$$

Для таксономии T_S можно построить такую РПС, которая будет унивалентной ей:

$$T_S \xrightarrow{G_\psi} \psi_S$$

$$T_S \cong \psi_S$$

Для решения задачи идентификации информации в тексте необходимо представить классификационные атрибуты, содержащиеся в онтологии, в форме, пригодной для работы с методом рекурсивной редукции. Одной из таких форм может быть активная онтология, которая имеет вид:

$$O_t = \langle X, R, F, A, D, R_s \rangle$$

где X – множество концептов;

R – конечное множество семантически значимых отношений между концептами;

F – конечное множество функций интерпретации, заданных на концептах и / или отношениях (отметка, какую информацию считывать);

A – конечное множество аксиом, которые используются для записи всегда истинных высказываний (определений и ограничений) в терминах тематики ПдО;

D – множество дополнительных определений понятий в терминах тематики ПдО (текстовые описания);

R_s – множество ограничений, определяющих область действия понятийных структур определенной тематики ПдО.

Для представления онтологического реестра в такой форме следует интерпретировать множество атрибутов следующим образом:

$$O_{st} = \langle X, R, F(A'), A(A'), D(A'), R_s(A') \rangle$$

Для этого необходимо построить G_t – преобразование интерпретации атрибутов:

$$O_S \xrightarrow{G_{O_t}} O_{St}$$

При этом:

- Конечные множества функций интерпретации F и дополнительных ограничений R_S создаются с помощью специализированной процедуры;
- Множество дополнительных определений понятий D может быть представлено с помощью определенных атрибутов без дополнительной интерпретации;
- Конечное множество аксиом A не используется в задаче классификации, поэтому можно принять $A = \emptyset$.

Активная онтология может быть преобразована в набор правил для рекурсивного редуктора. Каждое правило имеет следующую структуру:

$$g = \langle f_{ap}^g, f_{tr}^g \rangle,$$

где f_{ap}^g – функция применимости, определяющая, может ли правило быть применено к определенному набору входной информации;

f_{tr}^g – функция трансформации, задающая преобразование набора входной информации.

Правило такого вида может быть сформировано с помощью соответствующего преобразования:

$$R_S(A') \xrightarrow{G_{rd}} f_{ap}^g$$

$$F(A') \xrightarrow{G_{rd}} f_{tr}^g$$

где G_{rd} – преобразование формирования редуктивных правил.

Созданная таким образом база правил может быть использована для формирования специализированного оператора редукции, предназначенного для создания выколотых онтологий:

$$O_{nr} = \langle X, R, K(X) \rangle$$

где $K(X)$ – контексты объектов X , являющие собой подмножество атрибутов A' .

Выколотая онтология также может быть преобразована в РПС:

$$O_{nr} \xrightarrow{G_v} \psi_{nr}$$

Также дополнительно для решения задачи трансдисциплинарной интеграции информации РПС необходимо расширить контекстами:

$$W^T \rightarrow X_i(W^T)$$

где W^T – контекст;

X_i – концепт, который содержится в контексте.

Преобразование нарратива на принципах трансдисциплинарной интеграции формирует из "выколотых онтологий" "таксономию нарративного дискурса". Преобразование осуществляется с помощью функции поиска:

$$Q_s(H, T) = \{H \langle \{V(I) \times V(T)\} \rangle\}$$

где H – индекс, полученный в результате индексации массива нарративов сетевых документов с помощью специализированной функции индексации Q_H ;

$V(I), V(T)$ – идентификаторы лексемы I и таксономии T нарратива документа O_{nd} соответственно.

Функция поиска позволяет формировать связи между контекстами всех лексических единиц множества сетевых нарративных документов. Данная операция является основной для построения функции контекстной связности:

$$Q_C(T) = \bigcup_{x \in T} (Q_S(Q_H(C), L_x))$$

где C – множество нарративов сетевых документов O_{nd} , которые определяют содержательность информационной среды, в рамках которого осуществляется связывание;

T – таксономия, с лексическими единицами которой осуществляется связывание;

L_x – текстовое представление контекста лексической единицы x , принадлежащей таксономии T .

Множество индексов $\{H\}$, которое формируется на основе применения к множеству C функции индексации Q_H , формирует индексную зону \tilde{H} всех нарративов сетевых документов O_{nd} :

$$\tilde{H} = \{\{H\} \times T\}$$

Таким образом, функцию контекстной связности можно представить в виде:

$$Q_C(I) = \tilde{H} \times C \times \tilde{T} \times \mathfrak{R}^3$$

Функция контекстной связности создает условия для формирования нарративного дискурса на основе семантико-лексической и концептографической обработки всех нарративов сетевых документов O_{nd} .

Модель поведения системы

Модель поведения системы приводится для представления взаимодействий, отношений и зависимостей компонентов, из которых состоит система структурирования разностилевых ИР.

Модель поведения системы может быть представлена структурой:

$$M_D = \langle d_{use}, d_{act}, d_{seq} \rangle$$

где d_{use} – UML-диаграмма вариантов использования системы;

d_{act} – UML-диаграмма активности системы;

d_{seq} – UML-диаграмма взаимодействия системы.

UML-диаграмму вариантов использования программной системы структуризации разностилевых документов представлено на рис. 2.



Рисунок 2 Варианты использования системы

Данная диаграмма отражает участников процесса структурирования разностилевых документов и их основные возможности. Так, в рамках системы могут действовать «Пользователь», «Администратор» и

«Разработчик». Пользователь осуществляет классификацию текста по стилю, применяет рекурсивную редукцию для получения онтологий, просматривает информацию, а также формирует трансдисциплинарную интеграцию ИР. Администратор системы формирует и наполняет онтологический реестр стилей, формирует онтологические шаблоны обработки и осуществляет индексацию ИР. Разработчик, имеющий доступ ко всем функциям системы, создает онтологические шаблоны представления.

Алгоритм работы с системой структурирования документов различных стилей показывает диаграмма активности (рис. 3).

Пользователь может применять редукцию текстов, осуществлять их классификацию и пользоваться режимом трансдисциплинарного представления. В режиме администрирования создаются шаблоны представления и обработки, а также осуществляется анализ существующих и необходимых стилей.

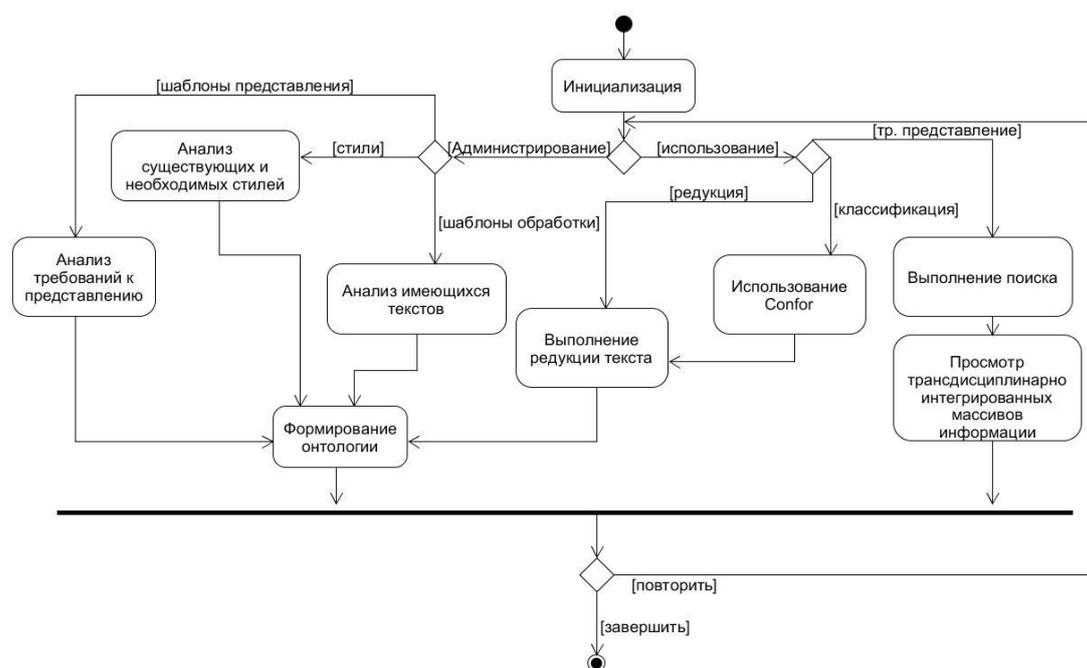


Рисунок 3 Диаграмма активности системы

На рис. 4 представлена диаграмма взаимодействия.

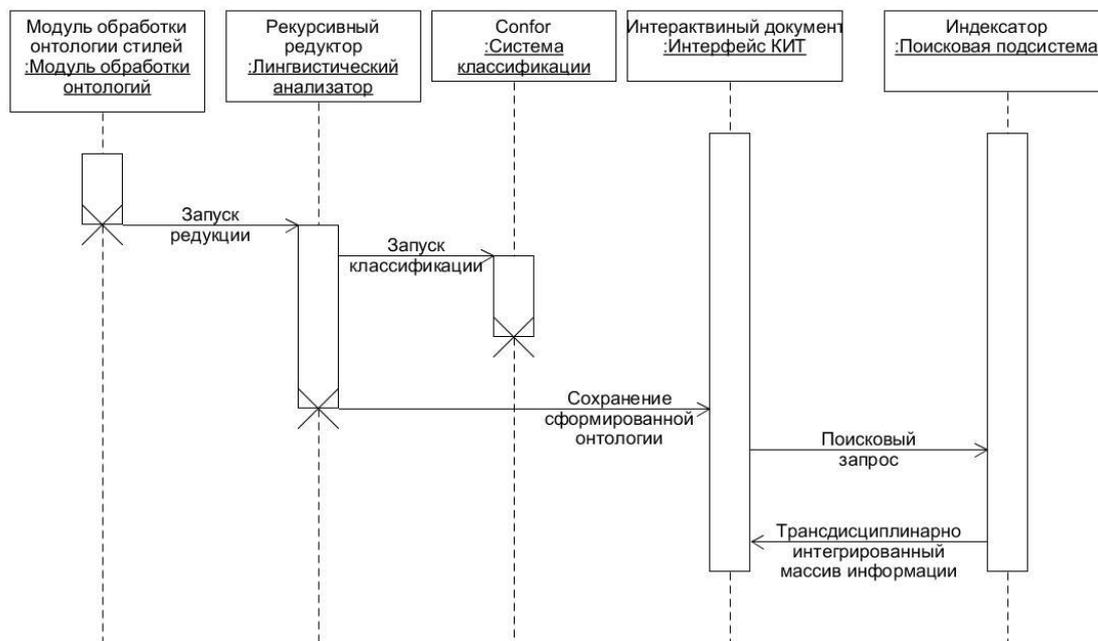


Рисунок 4 Диаграмма взаимодействия системы

Взаимодействие подсистем происходит в следующей последовательности:

- 1) В модуле обработки онтологий происходит обработка онтологии стилей;
- 2) Запускается редукция ЕЯТ к которой подключается классификация текста обрабатываемого по стилю;
- 3) Сформированная онтология отображается интерфейсом КИТ «Полиэдр» в виде интерактивного документа;
- 4) Элементы полученного интерактивного документа могут выступать в качестве поискового запроса, выполнив который с помощью индексации, система предоставляет пользователю трансдисциплинарно интегрированный массив информации.

Архитектура системы

Для обеспечения надежной и быстрой работы системы трансдисциплинарной интеграции ЕЯТ различных стилей используется трехуровневая архитектура, включая уровень представления, уровень логики и уровень данных (рис. 5).

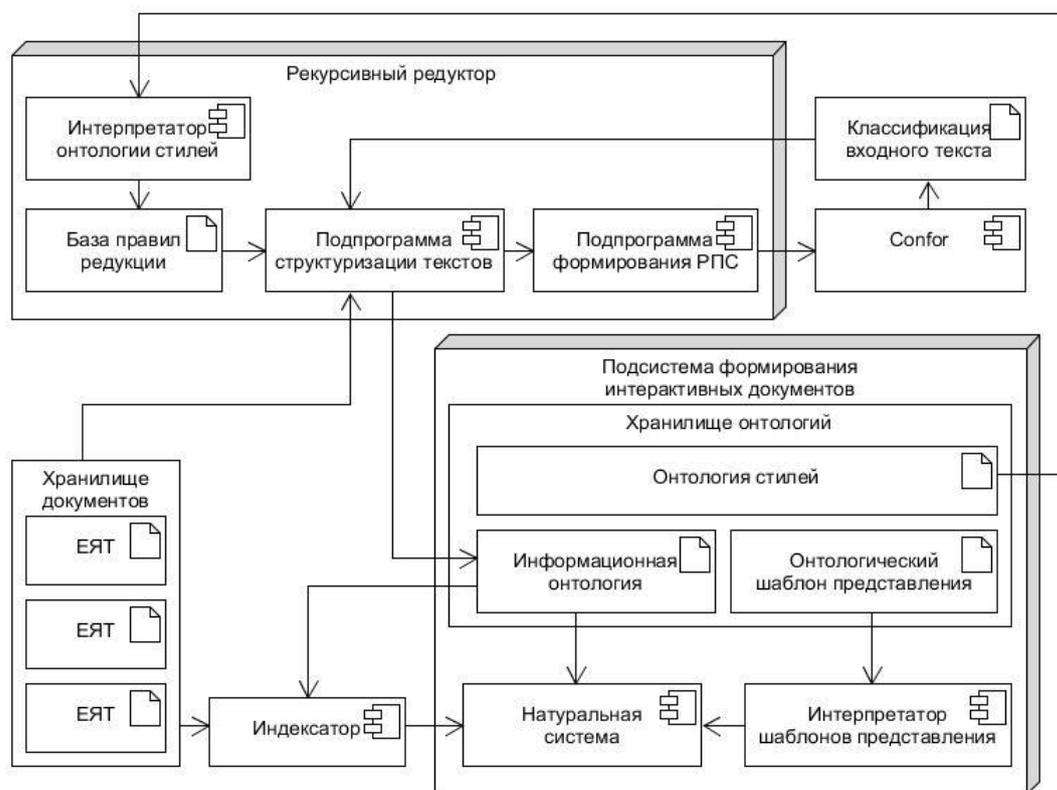


Рисунок 5 Базовая архитектура системы структурирования документов различных стилей

Система состоит из двух подсистем: подсистемы, осуществляющей структурирование текстов («Рекурсивный редуктор»), и подсистемы, обеспечивающей пользовательский интерфейс («Подсистема формирования интерактивных документов»).

Система предназначена для работы с распределенным массивом ЕЯТ, которые могут быть получены из различных источников. Центральным элементом системы является онтология стилей, содержащаяся в хранилище онтологий в рамках подсистемы формирования интерактивных документов. Данная подсистема в том числе обеспечивает интерфейс администратора для изменения онтологии стилей.

При запуске подсистемы структуризации онтология стилей извлекается из хранилища, и с помощью специализированной подпрограммы («интерпретатор онтологии стилей») превращается в подмножество базы правил редукции.

С учетом этих правил и осуществляется структуризация на первом этапе. Результаты структуризации превращаются в РПС с помощью соответствующей подпрограммы, и на их основе проводится классификация исходного текста с помощью классификационной системы «Confor». Полученная классификация подается на вход подпрограммы структуризации, что позволяет повторить редукцию с большей точностью и получить конечный результат - информационную онтологию, которая в дальнейшем хранится в соответствующей базе.

Подсистема формирования интерактивных документов содержит две основные составляющие: хранилище онтологий и натуральную систему. Дополнительно используется интерпретатор онтологических шаблонов представления, что позволяет модифицировать функциональность натуральной системы. Онтологические шаблоны представления, используемые в работе интерпретатора, хранятся в хранилище рядом с информационными онтологиями.

Дополнительно натуральная система позволяет выполнять трансдисциплинарную интеграцию информации. Для этого используется специализированный модуль индексации («Индексатор»). На вход индексатора подаются как структурированные (информационные онтологии), так и неструктурированные (ЕЯТ) документы, а результаты его работы отображаются с помощью натуральной системы.

При этом каждая из подсистем имеет свои уровни логики, данных и представления. Для подсистемы формирования интерактивных документов уровень данных обеспечивается хранилищем онтологий, а уровни логики и представления – натуральной системой. Подсистема структуризации не предусмотрена для прямого взаимодействия с пользователем и не имеет уровня представления, уровень логики обеспечивается подпрограммой структуризации, а уровень данных – специализированными подпрограммами считывания и записи файлов (в частности, интерпретатором онтологии стилей и подпрограммой формирования РПС).

Такая архитектура позволяет обеспечить максимально эффективное взаимодействие различных частей системы между собой и системы в целом – с пользователем.

Выводы

Структуризация ЕЯТ позволяет значительно повысить эффективность работы с имеющейся в них информацией. Разработка системы, обеспечивающей автоматизированную структуризацию, дает возможность эффективной работы с очень большими массивами данных, которые невозможно обработать вручную.

Применение онтологического классификатора стилей и двухуровневой схемы обработки (с дополнительной классификацией) позволяет значительно повысить точность выходного результата и уменьшить количество ошибок обработки.

Осуществление трансдисциплинарной интеграции позволяет в дальнейшем повысить эффективность обработки больших массивов разнородных пространственно распределенных документов.

Литература

- [Amit, 2005] Amit Konar. Cognitive Engineering: A Distributed Approach to Machine Intelligence. Series: Advanced Information and Knowledge Processing, Springer, Cham, 2005, 354 p.
- [David, 2013] David K. E. Modeling Narrative Discourse: Ph.D. thesis / David K. E. Columbia University, New York City, 2012, 383 p.
- [Malyshevsky, 1998] Malyshevsky A. Qualitative models in the theory of complex systems. In: Science. Fizmatlit, Moscow, 1998, 528 p.
- [Mayer-Schönberger, 2013] Mayer-Schönberger V, Cukier K. Big Data: A Revolution That Will Transform How We Live, Work, and Think. MA: Houghton Mifflin Harcourt, Boston, 2013, 252 p.
- [Басюк, 2017] Басюк Т. М., Досин Д. Г. Онтологічний інжиніринг. Вид-во Львівської політехніки, Львів, 2017, 224 с.
- [Величко, 2017] Величко В. Ю., Попова М. А., Приходнюк В. В., Стрижак О. Є. ТОДОС – ІТ-платформа формування трансдисциплінарних інформаційних середовищ. Системи озброєння і військова техніка, № 1, 2017, С. 10–19. http://nbuv.gov.ua/UJRN/soivt_2017_1_4.
- [Гладун, 2004] Гладун В. П., Ващенко Н. Д., Величко В. Ю., Ткаченко Ю. Г. Структуризация и анализ данных в растущих пирамидальных сетях. Syst. Res. Inf. Technol., no. 1, 2004, pp. 82–92. doi: 10.1017/SBO9781107415324.004.
- [Палагин, 2012] Палагин А. В., Крывый С. Л., Петренко М. Г. Онтологические методы и средства обработки предметных знаний. Изд-во ВНУ им. В. Даля, Луганск, 2012, 324 с.

- [Палагин, 2014] Палагин А. В. Трансдисциплинарность, информатика и развитие современной цивилизации. Вісник НАН України, no. 7, 2014, pp. 25–33.
- [Палагин, 2016] Палагин А. В. Онтологическая концепция информатизации научных исследований. Кибернетика и системный анализ. Т. 52, № 1, 2016, с. 3-9.
- [Приходнюк, 2018] Приходнюк В. В. Технологічні засоби трансдисциплінарного представлення геопросторової інформації [Текст]: дис. канд. наук.: (05.13.06 – Інформаційні технології). Інст. телеком. і глоб. інформ. Простору, Київ, 2018, 267 с.
- [Стрижак, Онтологические аспекты, 2014] Стрижак А. Е. Онтологические аспекты трансдисциплинарной интеграции информационных ресурсов. Открытые информационные и компьютерные интегрированные технологии, №. 65, 2014. С. 211–223.
- [Стрижак, Трансдисциплінарна інтеграція, 2014] Стрижак О. Є. Трансдисциплінарна інтеграція інформаційних ресурсів [Текст] : автореф. дис. д-ра техн. наук : 05.13.06. Нац. акад. наук України, Ін-т телекомунікацій і глобал. інформ. простору, Київ, 2014, 547 с
- [Стрижак, 2020] Стрижак О. Є. Таксономічні засади наративного дискурсу. Медична інформатика та інженерія. (In print)
- [Широков, 2017] Широков В. А. Язык. Информация. Система : Трансдисциплинарность в лингвистике. Palmarium Academic Publishing, Киев, 2017, 270 с.

Информация об авторах



Гайко Светлана Ивановна – аспирант, Институт телекоммуникаций и глобального информационного пространства НАН Украины, Киев-186, 03186, Чоколовский бульвар, 13; e-mail: svitgai@i.ua

Основные области научных исследований: обработка ЕЯТ, онтологический инжиниринг.



Приходнюк Виталий Валерьевич – кандидат технических наук, Национальный центр «Малая академия наук Украины», Киев, 04119, ул. Дегтяревская 38/44, 13; e-mail: Prikhodnyuk_Vitaly@nas.gov.ua

Основные области научных исследований: обработка ЕЯТ, онтологический инжиниринг.

MEANS OF TRANSDISCIPLINARY REPRESENTATION OF DIFFERENT STYLES INFORMATION RESOURCES

Svitlana Gaiko, Vitalii Prykhodniuk

Abstract: *The article deals with the problem of effective use the heterogeneous information resources (IR). A huge number of natural language documents accumulated on the web remain passive sources of knowledge. To work effectively with such documents, first of all, their structuring is required, which cannot be done manually, taking into account the necessary labor costs. It is also necessary to determine the way of presenting the results of IR structuring, since further possibilities when working with them depend on this. The most important of these capabilities is the integration of different styles IR that*

describe different subject areas in a single information space. The article proposes a model for the integration of heterogeneous documents that implements the principle of transdisciplinarity. The technology of automated processing of natural language texts (NLT) of various styles and the formation of interactive documents on their basis is described. The mechanism of displaying their structure, which is the narrative discourse, is considered. A mathematical model of this technology, a model of system behavior (in the form of UML diagrams), as well as the architecture of a software tool created on the basis of the described models are presented.

Keywords: *transdisciplinarity, ontology, growing pyramidal network (GPN), narrative discourse, functional styles.*