

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ТЕКСТОВОЙ ИНФОРМАЦИИ В ОБРАЗОВАТЕЛЬНОЙ ЭКСПЕРТНОЙ СИСТЕМЕ

Я.А. Ваграменко, Р.Г. Фанышев

Современные достижения в области информационных технологий позволили за короткий промежуток времени скопить в хранилищах данных различных организаций большие объемы информации, которая содержит скрытую информацию в виде знаний, поэтому задача аналитической обработки больших объемов информации становится весьма актуальной. Центральное место в процессах интеллектуального анализа естественно – языковой информации занимают следующие технологии: Data Mining, Text Mining и Semantic Web Ontology (язык OWL – Ontology Web Language), задачей которых является получение ранее неизвестных либо не выявленных знаний и закономерностей фактов в больших хранилищах данных. Источниками исходной информации для аналитической обработки могут являться базы знаний разных типов. Например, крупнейшая база знаний Internet, хранилища данных различных организаций. Значительная часть информации в этих источниках представлена в виде естественных текстов, процесс аналитической обработки которых требует создания принципиально новых моделей, методик и систем интеллектуального анализа информации [4].

Задача аналитической обработки естественных текстов является достаточно сложной и в общем случае связана с построением систем сложных интеллектуальных информационных систем. Однако необходимо взять во внимание тот факт, что информационно–аналитические структурные элементы и компоненты образовательного процесса гуманитарных вузов не нуждаются в извлечении всех закономерностей из естественных текстов в силу специфичности познавательной деятельности студентов. Вследствие чего можно сделать вывод о том, что необходимость построения модели естественного текста, реализующей глубинный семантический анализ текста,

не имеет смысла. Поэтому первоочередной задачей интеллектуального анализа текстовой информации в базах знаний образовательной экспертной системы является создание унифицированного формата мета-знаний, который бы позволил обеспечить необходимый уровень информативности и специфичности единиц знаний, адаптированных к международным стандартам в сфере технологий обучающих информационных систем [2].

На сегодняшний день для решения вышеперечисленных проблем представляется целесообразным использование технологий Semantic Web, потому что Semantic Web предполагает наличие у любой информации, находящейся в каком – либо хранилище, связанный с этой информацией точный смысл, который нельзя было бы перепутать даже в случае совпадения фраз или слов, встреченных в разных контекстах. Фактически это означает, что любая информация связывается с некоторым неотделимым от нее контекстом. Речь идет о двух подходах к извлечению необходимых фактов из естественно – языковых текстов.

Подход 1. Извлекается вся лингвистическая информация (синтаксис, анафорические связи), а затем на ее основе извлекаются факты.

Подход 2. Ищутся ключевые слова, которые наращиваются в тексте с помощью лингвистических правил до цепочки, описывающей факт.

В результате применения обоих подходов текстовый анализ сводится к формированию готовых словарей, способных в автоматическом режиме самообновляться [1].

В рамках учебных процессов применение web - онтологий позволит специфицировать основные компоненты учебных дисциплин - лекции, практики, лабораторные работы, используемые учебные материалы, а также обеспечит возможность организации эффективного распределенного доступа к учебным ресурсам, путем создания единой базы знаний, которая будет сочетать в себе множество учебных дисциплин и будет фактически распределенной по сети Интернет, что позволит сделать ее независимой от интерпретации конкретного учебного процесса. Роль обучающих систем в таком случае будет

сведена к роли интеллектуальных агентов, которые будут производить выборки из баз знаний в зависимости от контекста обучения (также возможно построения агентов для автоматического дополнения или изменения такой базы знаний новой информацией). Другой немаловажной особенностью такой системы – это возможность строить тестирующие программные системы, которые будут генерировать контрольные задания исходя из семантики описанных онтологий конкретных учебных курсов. Очевидно, что такие системы построения контроля знаний намного превосходят существующие на данный момент тесты, ориентированные на выборку одного из нескольких вариантов ответов.

Согласно принципам Semantic Web, процесс создания электронных документов разбивается на две части:

- создание web - онтологии документа, содержащей некоторые термины, понятийные структуры;
- визуализация содержимого онтологии, т.е. получение содержимого онтологии в некотором виде и формате [3].

Таким образом, в web-онтологии определяется смысл используемых понятий, характерных для конкретной дисциплины, т.е. специфицируются объекты предметной области, а с помощью языков трансформаций и форматирования – XSLT и XSL-FO получается визуальное представление содержимого онтологии в необходимом формате, например HTML, DOC и т.п.

Язык трансформаций XSLT позволяет выполнять трансформации структурированных документов, написанных на XML-подобных языках, например, OWL. Результатом трансформаций является некоторый набор данных, форматирование которого можно осуществить с помощью XSLT-FO.

Язык форматирования XSL-FO позволяет с большой точностью задавать макет и другую стилевую информацию, относящуюся к содержимому документов. Учитывая все вышеперечисленные принципы и возможности, был разработан способ стандартизации элементов образовательного содержимого учебных материалов, так называемого каркаса, для организации электронных

материалов учебных курсов с возможностью их последующего вывода как на экран, так и на печать.

Данный способ представляет собой шаблон, описывающий структуру электронных материалов учебного курса. Другими словами, мы создали онтологию, которая специфицирует структуру и понятия характерные для большинства создаваемых учебных курсов.

Предметной областью для нас является вся терминология используемая для организации учебного курса: тема, лекция, практическое занятие, лабораторная работа, контрольные вопросы, примеры, списки дополнительной литературы, а также все более мелкие компоненты каждого из объектов.

Так как, большая часть самостоятельной научно – познавательной деятельности студентов происходит во всемирно известном хранилище знаний Internet и подобных сетевых базах знаний (Intranet и корпоративно-вычислительные системы вузов), для реализации задач диссертационного исследования была выбрана технология формирования базы знаний образовательной экспертной системы на основе Web – онтологии «Учебная дисциплина». Разработанная Web – онтология «Учебная дисциплина» состоит из компонентов учебных дисциплин и материалов. В основу web-онтологии «Учебная дисциплина» были положены основные принципы, используемые для структуризации лекционных материалов, практических, лабораторных, курсовых и дипломных заданий. В соответствии с этими принципами была сформирована структурная схема Web онтологии «Учебная дисциплина». Ниже приведем описание данной онтологии с использованием технологии Semantic Web на языке C# реализующего механизм вывода Эйлера. Для того, чтобы разобраться с подходами, которые входят в состав специализированной библиотеки SemWeb на C#, мы воспользовались онтологией моделирующей простой конечный автомат. Ниже представлен исходный код описание некоторых функций:

```

internal class Program
{
    private static readonly string ontologyLocation =
        @»C:/dev/prototypes/semantic-web/ontologies/20074/states/»;
    private static string baseUri =
        @»file:///C:/dev/prototypes/semantic-
web/ontologies/2007/04/states/states.rdf#»;
    private static MemoryStore store = new MemoryStore();
    private static Entity Machine1 = new Entity(baseUri + «Machine1»);
    private static Entity Input1 = new Entity(baseUri + «I1»);
    private static Entity Input2 = new Entity(baseUri + «I2»);
    private static Entity theSystem = new Entity(baseUri + «This»);
    private static string hasInput = baseUri + «hasInput»;
    private static string isInState = baseUri + «isInState»;
    private static void Main(string[] args)
    {
        InitialiseStore();
        DisplayCurrentStates();
        SetNewInput(Input2);
        DisplayCurrentStates();
    }
    private static void DisplayCurrentStates()
    {
        SelectResult ra = store.Select(new Statement(Machine1, new
Entity(isInState), null));
        Debug.Write(«Current states: «);
        foreach (Statement resource in ra.ToArray())
        {
            Debug.Write(resource.Object.Uri);
        }
    }
}

```

```

    Debug.WriteLine(«««);
}
private static void InitialiseStore()
{
    string statesLocation = Path.Combine(ontologyLocation, «states.n3»);
    string rulesLocation = Path.Combine(ontologyLocation, «rules.n3»);
    Euler engine = new Euler(new N3Reader(File.OpenText(rulesLocation)));
    store.Import(new N3Reader(File.OpenText(statesLocation)));
    store.AddReasoner(engine);
}
private static void SetNewInput(Entity newInput)
{
    Resource[] currentInput = store.SelectObjects(theSystem, hasInput);
    Statement input = new Statement(theSystem, hasInput, Input1);
    store.Remove(new Statement(theSystem, hasInput, currentInput[0]));
    store.Add(new Statement(theSystem, hasInput, newInput));
    Resource[] subsequentState = store.SelectObjects(Machine1, isInState);
    Statement newState = new Statement(Machine1, isInState,
subsequentState[0]);
    store.Replace(new Statement(Machine1, isInState, null), newState);
}
}
}

```

Объединив данную онтологию с онтологией конкретной дисциплины, т.е. конкретными данными, соответствующими учебной дисциплине, мы получили полноценную информационную базу, с которой можно проводить различные действия, например, применив к данной онтологии инструкции визуализации XSL и XSL-FO, мы можем получить на выходе различные представления учебных материалов.

В контексте разработки онтологии мы рассматривали учебные материалы как некоторое упорядоченное множество экземпляров объектов предметной

области, отобранных по некоторому критерию, визуализация средствами XSLT которых давала возможность получать разнообразные представления информации в различных форматах, например: гипертекстовая страница или набор таких страниц (HTML); документ XML; документ формата PDF.

Таким образом, разработав онтологию «Учебная дисциплина» и заполнив ее конкретными данными, соответствующей учебной дисциплине, мы получили эффективное средство для поддержки интеграции существующих баз знаний в базу знаний экспертной системы информационной поддержки научно познавательной деятельности студентов.

Литература

1. Вохминцев А.В., Мельников А.В. Модель знаний на основе нечетких семантических гиперсетей для представления отношений между объектами в естественном тексте // Интеллектика, логистика и системология. 2002. Вып. 7. С. 21-33.

2. Применение Web-онтологий в задачах дистанционного образования. URL: <http://shcherbak.net/dist/>

3. Рыбина Г.В. Теория и технология построения интегрированных экспертных систем. М.: ООО Издательство «Научиздатлит», 2008. 482 с.

4. Novak J. Clarify with Concept Maps // The Science Teacher. NJ. 1991. № 58. С. 45-49.