

Черепанова Ю.Ю.

Харьковский национальный университет радиоэлектроники, Украина

Проверка знаний с ответами на естественном языке с использованием тезауруса семантических полей

Эффективность любой системы образования во многом определяется качеством средств проверки знаний обучаемого. Контроль необходим как при предварительном тестировании для определения начального уровня знаний, так и после прохождения курса обучения для контроля усвоения материала. В связи с возрастанием роли дистанционного обучения задача повышения эффективности и интеллектуальности средств анализа ответов стала особенно актуальной. Особенности дистанционного образования объясняются новыми требованиями к системам контроля.

Для более адекватной оценки знаний обучаемого возникает необходимость в нахождении новых методов тестирования знания теоретического материала с высокой реактивностью, простотой реализации и минимальными трудозатратами автора обучающего курса. Наилучшим методом представляется способ тестирования, при котором проводится анализ смыслового соответствия естественно-языкового ответа обучаемого и теории с целью установления степени его адекватности. При этом минимальными трудозатраты автора будут в том случае, если теория будет представлена в системе в виде естественно-языковых текстов либо модели, которую возможно построить по таким текстам с наиболее полной автоматизацией. Следовательно, возникает необходимость в механизме построения смысловой модели по естественно-языковому тексту. В настоящее время еще не существует полного решения данной задачи, но даже частное решение имеет реальный практический интерес.

Одним из возможных решений поставленной задачи может стать включение в состав системы контроля знаний тезауруса семантических полей [1] и использование его при тестировании для автоматизированного построения концептуальной модели текста на естественном языке. Результат тестирования

при этом определяется на основе сравнения полученной модели с концептуальными моделями предметной области или эталона ответа. Эталонная модель также может быть получена по неформализованному естественно-языковому тексту всего курса либо по тексту одного наиболее полного правильного ответа на конкретный вопрос. Таким образом, достоинством такого подхода является то, что эталонная модель строится автоматически, как и модель ответа тестируемого, что уменьшит трудозатраты автора. Эксперту по данной предметной области не обязательно разбираться в методах построения модели, и не надо перечислять все варианты ответов, достаточно дать один расширенный ответ на естественном языке. Несколько ответов может понадобиться ввести в том случае, если ответы на вопрос могут быть разными с концептуальной точки зрения. Такая ситуация может возникнуть в том случае, например, если проверяется знание понятия, которое имеет несколько верных толкований, которые могут использоваться в смежных областях знания.

Можно предложить несколько вариантов построения концептуальных моделей по тексту ответа, построенных с применением тезауруса семантических полей. В таблице 1 рассмотрены методы построения и анализа соответствия данных моделей.

Таблица 1. Методы построения концептуальных моделей текста.

Наименование метода	Построение модели	Анализ соответствия моделей
1. Метод множества	По тексту строится семантический объем	Находится пересечение семантических объемов ответа и эталона, при наличии в семантическом объеме ответа более 50% семантических множителей, входящих в семантический объем эталона, ответ считается удовлетворительным.
2. Метод линейного множества	По тексту строится семантический объем с приписыванием семантическим множителям последовательности	Рассчитывается коэффициент корреляции, учитывающий, кроме вхождения в семантический объемы семантических множителей, и порядок следования слов.

	встречаемости слова в тексте.	
3. Метод подстановки	Для каждого слова ответа ищется корреляционный заменитель (попадающий в то же парадигматическое семантическое поле) среди слов эталона, или наоборот.	Вводятся уровни соответствия моделей – если а) найдены все заменители; б) найдено более 50% заменителей; в) найдены все имеющие низкую частоту и высокую значимость в данной предметной области слова, - то ответ верен на определенном уровне.
4. Метод подстановки с учетом типа связи в предложении	При построении модели для каждого заменителя указывается тип связи с другими словами в ответе и эталоне.	Для анализа вводится корреляционный коэффициент, учитывающий, кроме проверки наличия заменителя, и совпадение связи между словами.

Семантический объем понятия представляет собой набор семантических множителей, соответствующих содержанию этого понятия. Понятие может быть представлено как словом, так и словосочетанием.

Методы и алгоритмы построения семантического объема и кодирования слов в семантические множители рассмотрены в [2].

Под тезаурусом семантических полей не имеется в виду фиксированный дескрипторный тезаурус, аналогичный применяемым в информационно-поисковых системах или в системах автоматического индексирования. Для возможности использования тезауруса семантических полей в системе контроля знаний к нему предъявляются следующие требования:

- тезаурус семантических полей должен содержать информацию, необходимую для построения концептуальных моделей текста;

- для выполнения требования упрощения наполнения системы контроля знаний должны существовать методы автоматизированного построения тезауруса семантических полей.

Для реализации предложенных методов проверки правильности ответов, тезаурус семантических полей должен содержать следующую информацию :

1. объекты:

– перечень слов представляющий собой словарь тезауруса (СТ) ;

- словоформы, соответствующие этим словам, которые встречаются в обрабатываемых текстах представляющий собой словник тезауруса (СлТ);
- определения слов, входящих в СТ в текстовой форме (О) ;
- перечень семантических множителей – словарь множителей тезауруса (СМТ);
- словарь устойчивых терминов (СТТ).

2. отношения, фиксируемые в тезаурусе:

- связи между СТ и СлТ, отражающие, какая словоформа определенного слова появилась в обрабатываемых текстах (Отношение "слово-словоформа" - ОССл) – словарь словоупотреблений (СС);
- связи между отношениями ОССл, отражающие наличие синтагматической связи между словами, в определенных формах, встретившиеся в текстах (отношение "ОСС") – словарь синтагматических ссылок (ССС);
- связь между словом и определением, показывающая наличие дефиниции данного слова, заполняющийся из толковых и терминологических словарей – словарь дефиниций (СД);

Подробнее схема тезауруса и алгоритм его наполнения представлены в [3]

При такой организации тезауруса, синтагматические связи фиксируются в словаре, а связи между словами и кодами, и парадигматические связи между словами могут быть получены алгоритмически, поэтому нецелесообразно хранить их в тезаурусе постоянно [4,5]. Алгоритмическое получение данных связей по необходимости имеет преимущество возможной модификации критериев выделения связей в зависимости от цели их получения (например, изменение частотных критериев при получении парадигматического поля, уровня компонентного анализа при построении семантического объема), что делает тезаурус более универсальным, расширяя сферу его применения.

Предложенные методы позволят решить некоторые проблемы повышения эффективности тестирования теоретических знаний в системах контроля знаний за счет уменьшения трудозатрат автора при наполнении системы. Для

эффективного применения предложенных методов необходимо совершенствование методов построения семантических полей, в частности комплексного поля, и разработка эффективных механизмов анализа соответствия концептуальных моделей.

Литература:

1. Павлов, П. Ф. О возможности построения тезауруса семантических полей и его применения в информационных системах [Текст] / П. Ф. Павлов, Ю. Ю. Черепанова, И. Ю. Шубин // Вісник ХДПУ. Збірка наукових праць. – Вип. 108. – Харьков: ХДПУ, 2000. – С. 41-46.
2. Черепанова, Ю.Ю. Методы и алгоритмы построения семантического объема слова [Текст] / Ю.Ю. Черепанова //Восточно-европейский журнал передовых технологий. Информационные технологии. – 4\2(34)2008. – С.21-25.
3. Черепанова, Ю.Ю. Контроль знаний с ответами на естественном языке [Текст] / Ю.Ю. Черепанова //Восточно-европейский журнал передовых технологий. Информационные технологии. – 4\2(40)2009. – С.32-36.
4. Черепанова, Ю.Ю. О теоретико-множественном и теоретико-категорном подходах к моделированию семантических полей [Текст] / Ю.Ю. Черепанова // Проблемы бионики: Всеукр. межвед. науч.-техн. сб. – 2001. – Вып 54. – С.75-78.
5. Павлов, П.Ф. О проблеме автоматизированного выявления семантических полей [Текст] / П.Ф. Павлов, О.И. Бабина, Ю.Ю. Черепанова // Вісник ХДПУ. Збірка наукових праць. – Вип. 42. – Харьков: ХДПУ, 1999. – С.81-85.