

АННОТАЦИЯ
диссертационной работы Смайыл Асель Маралбайқызы
на тему: «Модели и методы интеллектуальной информационно-
обучающей системы», представленной на соискание степени доктора
философии (PhD) по специальности
6D070300 – Информационные системы

Актуальность темы исследования: Актуальным является разработка образовательного ресурса, позволяющего организовать эффективный процесс обучения с учетом семантического представления ресурсов в рамках Смарт-Университета.

Опыт многолетней работы университетов и центров послевузовского образования показывает, что в системе непрерывного профессионального образования индивидуальное обучение является основным требованием, а главным ограничением в решении проблем индивидуального образования является время. Еще одним важным аспектом этого вопроса является содержание заявленных учебных программ, которые не учитывают индивидуальные потребности будущих специалистов. Процесс обучения одинаковый для каждого студента, материал для каждого курса также одинаковый. Для сильного пользователя содержание может быть слишком простым и недостаточным и, следовательно, неэффективным, а для слабого пользователя оно может быть сложным и непонятным. Таким образом, поиск альтернативных способов индивидуализации обучения по-прежнему остается актуальной проблемой. Качественно новые возможности для самообучения и совершенствования профессиональных знаний предоставляют новые информационные технологии дистанционного обучения с использованием локальных и глобальных сетей. В отличие от традиционно построенных курсов очного и тем более заочного обучения, использование информационных технологий открывает путь для обучения непосредственно на рабочем месте, что при правильной организации позволяет индивидуализировать процесс и выделять необходимое количество времени для обучения пользователей без каких-либо ощутимых перерывов в работе. Концепция компьютерного обучения основана на принципах автономии (самоуправления) процесса. Его реализация предполагает новое сочетание обучающих и контрольных программ с развитым компонентом взаимной моральной ответственности преподавателей и пользователей. Автономия в учебном процессе подразумевает не только независимость учебного заведения, но и право студента выбирать индивидуальный путь обучения в рамках многоуровневой системы образования.

Цель исследования. Разработка системы с использованием моделей и методов интеллектуальной информационной системы, основанной на семантическом представлении данных, с учетом особенностей траектории и индивидуализации процесса обучения.

Задачи исследования. Для достижения цели исследования были поставлены и решены следующие задачи:

- 1) Обзор и анализ существующих информационных систем для образовательных ресурсов;
- 2) Модели и методы построения образовательного ресурса;
 - 2.1) Формирование содержания образовательного ресурса;
 - 2.2) Составление контента на основе семантического анализа;
 - 2.3) Определение последовательности содержания;
 - 2.4) Управление количеством обучающих элементов в соответствии с Microlearning;
- 3) Разработка системы в виде веб-портала;
- 4) Тестирование программного обеспечения;
- 5) Внедрение программного обеспечения в Astana IT University.

Основные положения диссертации, выносимые на защиту:

- 1) Методология построения образовательного ресурса основана на семантическом анализе;
- 2) Установление коэффициента семантической близости пары тем на основе предложенной гипотезы;
- 3) Расположение выражений в контенте (на слайдах, текстах) определяется методом Microlearning;
- 4) Результат экспериментального исследования разработанной системы, показывающий повышение качества образования и эффективности за счет семантического представления ресурсов, также создает условия для цифровизации образовательного процесса.

Научная новизна работы:

Научно обоснованы использование моделей и методов формирования контента, которые предназначены для любого студента, учитывают особенности пользователя.

Впервые использованы методы семантического анализа для создания контента обучения.

Выдвигается гипотеза с использованием методов определения семантической близости слов с помощью контекстного множества, впоследствии эксперимента научная гипотеза была доказана с помощью метода корреляции Пирсона для определения последовательности контента, также в научной работе приведены правила для дистанционного обучения с помощью онтологии.

Доказано, что использование разработанной системы повышает качество образования и показывает эффективность.

Практическая значимость работы:

Теоретические и прикладные результаты, полученные в работе, могут быть использованы при проектировании и разработке системы для информационных и обучающих организаций. А также, полученные научные результаты могут быть разработаны как теоретическая и прикладная основа для автоматизации процесса обучения с учетом индивидуальных особенностей каждого студента (интеллектуализация).

Разработанная система с использованием моделей и методов интеллектуальной информационно-обучающей системы внедрена в

университете ТОО "Astana IT University", которая в настоящее время активно используется в обучении.

Методы исследования:

Семантический анализ (семантическая близость пары тем), алгоритм декомпозиции уровней, локальные ссылочные методы, такие как мера Дайса, индекс Джаккара и мера Кульчинского, а также метод корреляции Пирсона, методы Microlearning, непараметрический метод статистического анализа с использованием критерия Манна-Уитни.

Апробация и внедрения результатов исследования

Основные положения и научные результаты работы были доложены и обсуждены на семинарах кафедры "Информационные системы" Международного университета информационных технологий и на международных научно-практических конференциях: V Всероссийском конгрессе молодых ученых в Национальном исследовательском университете ИТМО (Россия, Санкт-Петербург, 2016); Международной научно - практической конференции "ИННОВАЦИИ В ОБРАЗОВАНИИ И НАУКЕ", посвященной 25-летию Независимости Республики Казахстан и 20-летию Университета Сулеймана Демиреля (Казахстан, Каскелен, 2016); Международной научно-практической конференции "Путь к независимости через десятилетия", посвященной 25-летию Независимости Казахстана (Казахстан, Алматы, МУИТ, 2016); 2-ой Международной конференции "Информационные технологии в науке и промышленности 2016" (Казахстан, Алматы, МУИТ, 2016); Международной научно – практической конференции "5-я Международная конференция Наука и общество – Методы и проблемы практического применения" (Канада, Ванкувер, 2018); в научно-практической ONLINE конференции «Цифровой Казахстан - внедрение механизма IT-технологий в Техническое и Профессиональное Образование: особенности использования модульных и кредитных технологий обучения в образовательном процессе» организованным Республиканским учебно-методическим объединением (Казахстан, Алматы, 2018); "3-ей Международной конференции по надежности и безопасности систем (ICSRS 2018)" (Испания, Барселона, 2018); 5-ой Международной конференции "Цифровые технологии в науке и промышленности-2019" (Казахстан, Алматы, 2019); 2021 IEEE Интеллектуальные информационные системы и технологии (SIST) (Казахстан, Нур-Султан, 2021) и Международной научно-исследовательской конференции в Дортмунде 2021 (IRC 2021).

Объем и структура диссертации

Структура работы определяется поставленной целью и последовательностью решения сформулированных задач и строится по проблемно-тематическому принципу. Диссертация состоит из введения, четырех глав, заключения и списка литературы, включающего 77 наименований использованных источников.

Публикации по теме диссертации

Основные результаты, полученные в ходе диссертационной работы, опубликованы в 12 печатных работах, из которых 4 статьи опубликованы в

изданиях, рекомендованных Комитетом по контролю в сфере образования и науки Министерства образования и науки Республики Казахстан, 1 статья опубликована в издании, индексируемом базами данных Scopus и Web of Science Core Collection, с импакт-фактором 1,635, квартиль для компьютерных наук и информационных систем-Q3, процентиль для приложений в области компьютерных наук – 51, 4 статьи опубликованы в сборниках международных зарубежных конференций (Канада, Испания, Украина, Германия), 6 статей опубликованы в сборниках международных научно-практических конференций (Казахстан) и 1 авторское свидетельство о внесении сведений в государственный реестр прав на объекты, охраняемые авторским правом, № 19399 от 14 июля 2021 года на программу ЭВМ "Интеллектуальная информационная система с использованием моделей и методов семантического представления данных".

Во введении дается обоснование актуальности выбранной темы диссертации. Сформулированы: цель, объект, предмет и задачи диссертации. Описаны результаты проведенных исследований, показана их научная новизна и практическая значимость. Приведены данные об апробации основных результатов диссертационной работы.

В первом разделе диссертации рассматривается важность и роль интеллектуальной информационной системы для образовательных учреждений. Проведен обзор и анализ современных подходов к автоматизации образовательных процессов, а также обзор литературы. Рассмотрены особенности содержания дисциплины, а также последовательность и количество учебных элементов.

Информационные технологии занимают центральное место в процессе интеллектуализации общества, развития его системы образования и культуры. Кроме того, использование информационных средств обучения оказалось очень эффективным методом как для систем самообразования, так и для систем повышения квалификации и переподготовки кадров.

Процессы информатизации современного общества и тесно связанные с ними процессы информатизации всех форм образовательной деятельности характеризуются процессами совершенствования и массового распространения современных информационно-коммуникационных технологий (ИКТ). Такие технологии активно используются для передачи информации и обеспечения взаимодействия преподавателя и студента в современных системах открытого и дистанционного образования. Современный преподаватель должен не только обладать знаниями в области ИКТ, но и быть специалистом по их применению в своей профессиональной деятельности. Концепция технологии включает в себя применение научных и инженерных знаний для решения практической задачи. Тогда информационные технологии можно рассматривать как процесс превращения знаний в информационный ресурс. Целью информационных технологий является получение информации для ее последующего анализа и принятия на ее основе решения о выполнении того или иного действия. Информационно-коммуникационные технологии - это общее понятие, описывающее различные

устройства, механизмы, методы и алгоритмы обработки информации. С помощью сетевых средств ИКТ становится возможным иметь широкий доступ к учебно-методической и научной информации, организовывать оперативную консультационную помощь, моделировать исследовательскую деятельность, проводить виртуальные учебные занятия (семинары, лекции) в режиме реального времени.

Анализируя социальную значимость информационных технологий, представляется целесообразным констатировать следующее:

- Современные информационные и телекоммуникационные технологии позволяют активизировать и эффективно использовать информационные ресурсы общества, которые являются важнейшим стратегическим фактором его развития;

- Развитие цивилизации происходит в направлении информационного общества, в котором объектами и результатами труда большинства занятого населения являются уже не материальные ценности, а главным образом информация и научные знания. В то же время информационные технологии позволяют оптимизировать и, во многих случаях, автоматизировать информационные процессы, происходящие в обществе;

- Информационные процессы являются важными элементами других, более сложных производственных или социальных процессов. В связи с этим информационные технологии рассматриваются как составные части соответствующих производственных или общественных технологий;

- Телекоммуникационные технологии, являясь частью информационных технологий, играют чрезвычайно важную роль в обеспечении информационного взаимодействия между людьми и организациями, а также в системах подготовки и распространения массовой информации.

Было отобрано 7 Высших Учебных Заведений Казахстана, таких как: Astana IT University, Международный Университет Информационных Технологий, Университет Нархоз, AlmaU, UIB, Казахский Национальный Университет имени Аль-Фараби и Медицинский университет Астаны. С 16 марта 2020 года все университеты полностью перешли на дистанционное обучение в связи с пандемией во всем мире. Все вышеперечисленные университеты используют систему "Moodle", так как у системы есть преимущества: модульная веб-ориентированная структура, поддержка мультимедийных данных, есть возможность работать с курсами, созданными в соответствии с международными стандартами SCORM и IMS, а также система с открытым исходным кодом.

Анализ дистанционного образования выявил следующие преимущества: гибкий график учебного процесса, а также доступность материалов и мобильность. Недостатки: самодисциплина, ответственность, терпимость, целеустремленность и настойчивость.

Адаптивные интеллектуальные системы предполагают: гибкость обучения в интерактивной образовательной среде, персонализацию и адаптацию обучения, свободный доступ к контенту независимо от географии.

Такие технологии позволяют разрабатывать учебно-методические материалы, а также формировать индивидуальные траектории обучения.

Среди 94 респондентов был проведен опрос на русском языке об использовании адаптивной интеллектуальной системы обучения. Вопросы были посвящены дистанционному обучению. Так как адаптивная система в первую очередь может быть использована в дистанционном обучении. В одном из вопросов необходимо было указать, из какого города пользователь, заполняющий анкету, или пользователь хотел бы учиться. Результаты анкетирования показали, что респонденты хотят учиться с помощью адаптивной интеллектуальной системы поскольку, это сейчас актуально. Основываясь на результатах анкетирования, можно сделать вывод, что студенты не хотят тратить время на предметы, которые они уже знают, они предпочитают индивидуальную траекторию обучения. Система должна быть адаптивной и интеллектуальной для пользователей. Обзор и анализ показали, что проблема с последовательностью контента существует и не решена, в связи с этим в данной работе рассматриваются особенности взаимосвязи различий между материалами, определяется последовательность контента на основе семантического анализа данных, а также реализуется система с индивидуальным процессом обучения.

В традиционном обучении есть недостатки в формализации процессов, так как обучение одинаково для всех студентов. Контент один и тот же, а типы восприятия пользователей разные, также возможно, что пользователь знает весь курс, но он должен его пройти. Традиционная система обучения не является гибкой и не индивидуальной для каждого пользователя. Существующие методы формализации обучения являются слабыми, и желательно повысить эффективность обучения за счет индивидуализации обучения. Развитие системы с использованием методов и моделей повысит эффективность обучения. В разделе также упоминается многокритериальная модель обучения, которая также обеспечивает хорошую эффективность обучения. Обзор литературы показал, что в мире существует тенденция индивидуального обучения. Многие ученые разработали адаптивные системы для пользователей. Адаптивная интеллектуальная система поможет всем пользователям эффективно тратить время на обучение, получая при этом необходимые знания.

Во втором разделе диссертации рассматривается методология создания интеллектуальной информационно-обучающей системы. В частности, содержание процесса обучения должно зависеть от типа восприятия пользователя. А семантический контент-анализ с помощью онтологии обеспечит индивидуальную траекторию для студента.

В связи с информатизацией образования идеи адаптивного обучения были также использованы в компьютерном обучении. Основное требование, чтобы информационная система обучения была разработана с учетом принципов адаптивного обучения, заключается в обеспечении процесса обучения (как в рамках университета, так и при дистанционном обучении) в соответствии с индивидуальными особенностями обучающегося. Эта задача

может быть решена путем внедрения в систему обучения различных приемов и методов, связанных с различными вариантами функциональности системы обучения и различными способами ее реализации.

Психофизиологические способности по типу восприятия делятся на визуальные, аудиальные и кинестетические. Диагностика доминирующей перцептивной модальности С.Ефремцева - это методика ведущего канала восприятия. Визуалы воспринимают информацию через образы, память, воображение. Визуалы имеют повышенную жестикуляцию. Аудиалы воспринимают информацию через слух. Кинестетики воспринимают информацию через чувства, эмоции. Следует отметить, что человека, полностью принадлежащего к одному типу мышления, не существует. В каждом из нас есть частичка от каждого из них, но в то же время у человека есть доминирующий тип восприятия. Согласно статистике, у 35% людей доминирующим типом восприятия является визуальный, у 25% людей доминирующим типом восприятия является аудиальный, а у 40% людей доминирующим типом восприятия является кинестетический.

Методы визуализации контента для визуалов: диаграммы, графики, иллюстрации, демонстрационные модели, эксперименты и эксперименты. Для аудиалов: лекционные материалы, аудиокурсы, видео, аудиокниги.

В рамках курса “ Database in Information Systems ” в Международном Университете Информационных Технологий был проведен тест на определение типа мышления для 3 групп студентов по методу ведущего канала восприятия.

Интеллектуальная система образования может помочь пользователям эффективно учиться. Информационная система предполагает:

- гибкость обучения в интерактивной образовательной среде;
- персонализация и адаптация обучения;
- разнообразный контент по типу восприятия;
- бесплатный доступ к контенту независимо от географии.

Интеллектуальная система позволяет разрабатывать революционные учебные материалы, а также формировать индивидуальные траектории обучения.

Адаптивность системы заключается в том, что у каждого пользователя будет своя траектория обучения. При входе в систему пользователи проходят 2 теста, которые определяют тип восприятия и уровень знаний по конкретному курсу.

Интеллектуальные информационные системы являются естественным результатом развития традиционных информационных систем. Они сконцентрировали в себе самые высокотехнологичные технологии с высоким уровнем автоматизации не только процессов подготовки информации для принятия решений, но и процессов разработки решений на основе данных, полученных информационной системой.

Разработка онтологии включает в себя:

- 1) определение классов в онтологии;

2) расположение классов в таксономической иерархии (подкласс-суперкласс);

3) определение слотов и описание допустимых значений этих слотов;

4) заполнение значений слотов экземпляра.

16 классов с атрибутами созданы для системы: у класса «departments» есть id, name, created_at, updated_at; у класса «users» есть id, name, email, password, remember_token, created_at, updated_at; у класса «students» есть id, character_type, user_id, created_at, updated_at; у класса «teachers» есть id, name, user_id, department_id, created_at, updated_at; у класса «subjects» есть id, name, department_id, created_at, updated_at; у класса «teacher_subjects» есть id, teacher_id, subject_id, created_at, updated_at; у класса «quizzes» есть id, title, teacher_id, subject_id, isPsychological, created_at, updated_at; у класса «question_types» есть id, name, created_at, updated_at; у класса «student_quizzes» есть id, student_id, quiz_id, accepted, result, created_at, updated_at; у класса «questions» есть id, title, question_value, quiz_id, question_type_id, created_at, updated_at; у класса «themes» есть id, name, subject_id, created_at, updated_at; у класса «materials» есть id, title, created_at, updated_at; у класса «answers» есть id, right, content, question_id, created_at, updated_at; у класса «question_themes» есть id, question_id, theme_id, created_at, updated_at; у класса «theme_materials» есть id, material_id, theme_id, character_type, created_at, updated_at и у класса «student_quiz_results» есть id, question_id, answer_id.

Для методология создания контента используется онтологии. Разработка онтологии начинается с определения классов, затем необходимо распределить классы в таксономической иерархии, затем определение слотов и описание допустимых значений этих слотов и в конце заполнение значений слотов экземпляра. Таким образом, можно создать онтологию контента для образовательных учреждений. Также определяется распределение контента по типу восприятия пользователем. Создание онтологии поможет создать условия для индивидуализации обучения. Использование различных методов семантического анализа данных позволяет создать индивидуальную траекторию обучения для пользователей.

В третьем разделе диссертации рассматриваются различные методы и модели расчета семантической близости слов и тем. Задача решается с помощью последовательности тем в одной дисциплине, а также с помощью количества учебных элементов. Предложена гипотеза, а также доказательство гипотезы с использованием различных научных методов. Разработана математическая модель информационной системы. Также определяется оценка эффективности обучения с помощью интеллектуальной системы.

Рассмотрим алгоритм создания контента. Взаимосвязь между темами T_N любого курса может быть представлена в виде ориентированного графа, G_T вершины графа являются темами, а дуги определяют взаимосвязь между темами. Свойства графа темы G_T :

1) Существует начальная вершина, соответствующая началу процесса обучения, и конечная вершина, соответствующая концу процесса обучения.

2) G_T - это бесконечный граф, так как, освоив последовательности тем, началом которой является некоторая тема T_N , студент не может начать изучать ее снова.

3) Поскольку граф G_T бесконечен, его можно разложить на уровни, то есть представить в виде иерархии.

В этом случае на верхнем уровне будет фиктивная начальная вершина, а на нижнем уровне - фиктивная конечная вершина, которая соответствует окончанию процесса обучения. Итоговая вершина определяется на основе анализа компетенций, которыми должен обладать студент после окончания учебы.

Рассмотрим разложение алгоритма на уровни бесконечного графа G_T : 1) Найти вершину без входящих дуг и присвоите ей ранг $r = 0$. Удалить дуги из этой вершины.

2) Предположим, что на каком-то этапе в графе нет вершин без входящих дуг. Присвойте этим вершинам следующее значение ранга r и вычеркните дуги, выходящие из них.

3) Этап 2 повторяется до тех пор, пока все вершины не будут ранжированы.

Замечание 1. Свойство бесконечности является наследственным, то есть, если какая-либо вершина графа удаляется вместе с падающими дугами, оставшийся подграф также является бесконечным.

Замечание 2. Неконтурные графы обладают следующими свойствами: – Существует по крайней мере одна вершина без входящих дуг (начальная вершина);

– Существует по крайней мере одна вершина без исходящих дуг (называемая конечной);

– Граф может быть представлен в виде иерархии, то есть разложен на уровни, при этом номер уровня - это длина максимального пути от начальной вершины до вершины этого уровня.

Использовался метод расчета семантической близости профессора Бондарчука. Предположим, что w_1 и w_2 - это слова, для которых необходимо вычислить семантическую близость. Метод Бондарчука можно разделить на несколько этапов:

1. Формирование контекстных наборов слов w_1 и w_2 . Эти наборы содержат слова, с которыми слова w_1 и w_2 часто используются в одном контексте. Затем формируется общий контекстуальный набор слов. Очевидно, что мощность этого набора будет равна $n+m$.
2. Вычисление нормализованных свойств между общим определителем и каждым из слов w_1 и w_2 .
3. Вычисление семантической близости. Для расчета семантической близости нужно вычислить коэффициенты R_i для всех слов из набора контекстов C .

Также необходимо рассчитать семантическую близость всех тем друг с другом, показанную в таблице 1:

Таблица 1. Семантическая близость между темами

Topics	w_1	w_2	Semantic proximity
Topic 1 - Introduction to the Information Systems	Topic 1	Topic 2	0,3
Topic 2 - Software Requirement Specification (part 1)	Topic 2	Topic 3	0,8
Topic 3 - Software Requirement Specification (part 2)	Topic 3	Topic 4	0,77
Topic 4 - Information System modeling	Topic 4	Topic 5	0,483
Topic 5 - UML modeling	Topic 5	Topic 6	0,81
Topic 6 - Design of IS	Topic 6	Topic 7	0,44
Topic 7 - Database	Topic 7	Topic 8	0,21
Topic 8 - Search Engine	Topic 8	Topic 9	0,37
Topic 9 - Software Testing Information System	Topic 9	Topic 10	0,48
Topic 10 - Feasibility study of the IS	Topic 10	Topic 11	0,91
Topic 11 - Criteria for IT projects	Topic 11	Topic 12	0,27
Topic 12 - Development of IS	Topic 12	Topic 13	0,67
Topic 13 - ADO technology	Topic 13	Topic 14	0,34
Topic 14 - Final presentation, Topic 15 - Examination	Topic 14	Topic 15	0,94

Выдвигаемая гипотеза: если семантическая близость больше 0,75 (в результате экспертных опросов и эксперимента), то темы являются зависимыми, а именно определяется большая семантическая связь между темами. Другими словами, содержание этих тем взаимосвязано и зависит друг от друга. И, конечно, темы совпадают. Если придерживаться гипотезы, то зависимость проявляется в темах:

1. зависимость - T2, T3;
2. зависимость – T3, T4;
3. зависимость – T5, T6;
4. зависимость – T10, T11;
5. зависимость – T14, T15.

Правила:

Правило 1: Если студент не освоил 2-ю тему, он не сможет перейти к 3-й теме;

Правило 2: Если студент освоил тему 2, но не освоил тему 3, он не сможет перейти к теме 4;

Правило 3: Если студент не освоил тему 5, он не сможет перейти к теме 6;

Правило 4: Если студент не освоил тему 10, он не сможет перейти к теме 11;

Правило 5: Если студент не освоил тему 14, он не сможет перейти к теме 15.

Используя семантическое представление данных, а также семантический анализ текста, можно разработать последовательность тем в рамках одной темы. И в дальнейшем можно использовать этот метод не только в учебном процессе, но и в курсе, или при составлении плана на учебный год.

Научный эксперимент проводился в Astana IT University в течение первого триместра 2019 года среди студентов первого курса. В этом эксперименте приняли участие 115 пользователей, разделенных на 5 групп. Студенты сначала прошли тест на определение уровня знаний, затем тест на определение типа восприятия. Для образовательной системы существует только 2 типа восприятия: визуальное и аудиальное. Тип восприятия определялся для типа контента. Для визуалов элементов система предоставляла больше видео, презентаций, для аудиоматериалов - только аудиокнижки и лекции. После каждой лекции давались задания для выполнения, как показано в таблице 2. В течение первого триместра пользователи могут сдавать все экзамены, включая практические или лабораторные задания и тесты.

Таблица 2 – все оценки одной группы

№	Users	type	level kn	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15
1	1_student	audial	22	95	49	64	70	72	78	85	93	87,5	95	90	80	81	87	82,5
2	2_student	visual	20	95	63	73	55	74	76	85	85	75	85	80	70	74	75	75
3	3_student	visual	20	95	43	50	60	57	72	85	92	87,5	100	75	80	80	90	81,6
4	4_student	audial	23	70	43	53	65	32	53	85	85	37,5	75	70	50	62	74	64,1
5	5_student	audial	22	95	60	60	70	88	83	80	90	56	85	85	70	73	84	80,4
6	6_student	audial	16	95	76	73	69	75	85	90	90	75	75	77	60	75	100	88
7	7_student	visual	15	55	76	70	59	30	44	60	74	62,5	85	80	60	64	50	52,4
8	8_student	audial	15	95	76	65	50	90	84	70	85	87,5	100	90	70	75	100	87,7
9	9_student	audial	15	95	43	33	75	63	73	50	50	62,5	80	85	60	60	65	65,9
10	10_student	audial	7	0	66	70	83	71	75	80	75	0	85	80	60	63	77	72,2
11	11_student	audial	13	80	76	70	59	57	61	75	70	50	70	75	70	69	64	64,6
12	12_student	visual	17	85	76	64	60	85	85	90	90	87,5	90	85	85	83	85	84,4
13	13_student	audial	13	0	43	53	49	52	54	75	80	0	0	75	70	59	60	57,9
14	14_student	audial	21	95	46	57	83	78	77	60	90	0	100	75	55	90	99	89,7
15	15_student	visual	29	95	76	70	60	73	77	85	85	75	85	90	80	79	85	80,8
16	16_student	audial	17	85	60	53	70	73	78	85	90	87,5	90	90	80	79	79	78,7
17	17_student	visual	18	80	63	67	60	85	75	80	90	0	0	60	70	90	90	85,5
18	18_student	audial	28	85	49	50	66	93	83	70	84	87,5	85	85	85	78	80	80,3
19	19_student	audial	7	95	49	57	64	56	77	75	75	75	85	90	75	75	79	77,2
	Average	-	-	78,4	59,6	60,6		68,6	73,1		82,7		81,6					
				2	3	3	64,6	3	5	77,1	8	60,75	6	80,89	70	74,15	80,15	76,25

Доказательство гипотезы. Но - нулевая гипотеза, все темы не связаны между собой. На - альтернативная гипотеза, темы связаны, где семантическая близость больше 0,75. Используя локальные ссылочные методы, такие как мера Дайса, индекс Джаккара и мера Кульчинского, а также метод корреляции Пирсона, можно охарактеризовать существование линейной зависимости между двумя величинами. Преподаватели дали экспертную оценку взаимосвязи между темами каждой лекции. Нужно рассчитать семантическую близость, используя метод Бондарчука, индекс Джаккара, меру Кульчинского, меру Дайса, экспертную оценку и корреляцию Пирсона. Студенты сдавали задания, и средние значения были использованы для всех методов. Краткое изложение полного текста было взято из каждой темы. По количеству слов

между темами можно рассчитать близость слов, используя различные методы. Каждая тема была оценена, и итоговый балл был принят за среднее значение. Подставляя формулы, получили значения для всех методов. В таблице 3 показаны пары тем, а также рассчитана семантическая близость с использованием метода Бондарчука, индекса Джаккара, меры Кульчинского, меры Дайса, экспертной оценки и корреляции Пирсона.

Таблица 3 - Семантическая близость тем с использованием различных методов

w_1	w_2	Bondarchuk's semantic proximity	Jaccard' index	Kulchinsky' measure	Dice' measure	Expert evaluation	Pearson's correlation
Topic 1	Topic 2	0,3	0,1379	0,2244	0,2666	0,3	0,05246
Topic 2	Topic 3	0,8	0,0638	0,1428	0,0857	0,9666	0,77802
Topic 3	Topic 4	0,77	0,0862	0,1607	0,1785	0,75	0,7516
Topic 4	Topic 5	0,483	0,0416	0,0881	0,0909	0,4666	0,06282
Topic 5	Topic 6	0,81	0,025	0,049	0,0526	0,9	0,89231
Topic 6	Topic 7	0,44	0,0363	0,0789	0,0526	0,45	0,28259
Topic 7	Topic 8	0,21	0,0338	0,0697	0,0869	0,3166	0,68234
Topic 8	Topic 9	0,37	0,0172	0,0356	0,0277	0,3166	0,12597
Topic 9	Topic 10	0,48	0,0172	0,0356	0,0434	0,55	0,55783
Topic 10	Topic 11	0,91	0,0192	0,0384	0,0333	0,883	0,82699
Topic 11	Topic 12	0,27	0,0174	0,1333	0,0133	0,25	0,46811
Topic 12	Topic 13	0,67	0,0185	0,0366	0,04	0,5333	0,44746
Topic 13	Topic 14	0,34	0,0232	0,0463	0,0526	0,3666	0,73213
Topic 14	Topic 15	0,94	0,0243	0,048	0,0434	0,9166	0,95878

Корреляция и семантическая близость показывают, что они ближе к экспертной оценке. Таким образом, можно сделать вывод, что там, где семантическая близость превышает 0,75, темы являются зависимыми и согласованными. Для этого предмета существует 5 пар тем, которые взаимосвязаны, как описано в гипотезе. При семантической близости более 0,75 по методу Бондарчука и корреляции Пирсона наблюдается сходство данных. Во время тестирования пользователи использовали интеллектуальную систему для обучения. Благодаря системе получили полностью контент, основанный на типе восприятия. Студенты также сдавали тесты и задания. Таким образом, можно видеть начальный уровень знаний и конечные знания. Оценки за задания также аналогичны семантической близости тем. Семантический анализ, а также методы семантической близости тем помогают решить многие проблемы в образовательной сфере: последовательность тем, количество учебных элементов и содержание.

Для информационно-обучающей системы разработана математическая модель, которая включает в себя 4 различных коэффициента (1) и имеет условие (2).

$$W = \frac{1}{4}(\alpha w_1 + \beta w_2 + \gamma w_3 + \delta w_4) \quad (1)$$

где α - коэффициент семантической близости Бондарчука,
 β – коэффициент индекса Жаккара,
 γ - коэффициент меры Кульчинского,
 δ - коэффициент измерения Дайса.

Условие характеризует выдвинутую гипотезу: если любое слово $x(i)$ встречается в тексте $T1$, $y(j)$ встречается в тексте $T2$ и семантическая близость двух тем равна или превышает 0,75, то эти две темы похожи по смыслу, в противном случае две темы разные.

$$\begin{aligned} \text{If } \forall (x(i) \in T1) \text{ and } ((y(j) \in T2) \text{ and } (\rho(x(i), y(j))) \geq 0,75 \\ \Rightarrow (x(i), y(j) \subseteq z) = 1 \\ \text{else } (x(i), y(j) \not\subseteq z) = 0 \end{aligned} \quad (2)$$

где $x(i)$ – i слово в тексте $T1$,
 $y(j)$ – j слово в тексте $T2$,
 $\rho(\cdot)$ – расстояние, рассчитанное по методу Бондарчука,
 $\rho(\cdot) \geq 0,75$.

Микрообучение может помочь людям расширить свои знания и не тратить много времени впустую. Для создания содержания образования в *Microlearning* необходимо извлекать знания и использовать аннотации, тем самым сжимая основные факты и получая необходимую информацию.

По данным Microsoft, средняя продолжительность концентрации внимания сейчас составляет всего 8 минут. Микрообучение должно содержать контент, особенно продолжительность видео или аудиоматериалов должна составлять 5-10 минут, количество слайдов в презентациях должно составлять 5-10, а количество страниц в текстовых материалах должно составлять 1-2 страницы, тестовые и практические задания. Структура *microlearning* аналогична структуре традиционного метода обучения, но есть различия, такие как тестирование начального уровня знаний, урок (аудио/видео и текстовый материал), тестовые или практические задания и итоговое тестовое/практическое задание. У каждой категории пользователей есть свой собственный мини-курс. Общее количество категорий составляет - 10.

Рассматриваются в работе алгоритмы, модели и методы создания интеллектуальной информационной системы для образования. Алгоритм декомпозиции уровня используется для решения последовательности содержимого. Выдвигается гипотеза и даются доказательства с использованием различных методов, включая расчет семантической близости метода Бондарчука, индекса Джаккара, меры Кульчинского, меры Дайса, экспертной оценки и корреляции Пирсона. Предложена математическая модель системы обучения, основанная на семантической близости тем. Модель используется для извлечения знаний для метода *microlearning*. *Microlearning* позволяет пользователям быстрее и проще воспринимать и усваивать информацию. Рассчитана оценка эффективности использования системы непараметрическим методом статистического анализа с

использованием критерия Мана-Уитни. Эффективность обучения с использованием системы в 1,15 раза выше.

В четвертом разделе описывается техническая реализация интеллектуальной информационно-обучающей системы “edu-elt.kz”. Программное обеспечение реализовано в виде веб-сайта с использованием языка программирования общего назначения PHP с открытым исходным кодом, СУБД интеграции MySQL и веб-платформы Laravel, предназначенной для разработки с использованием архитектурной модели MVC. Выбор языка программирования PHP обусловлен тем, что это язык сценариев, созданный для сайтов с динамическим контентом. Пользователь может видеть функции курса: продолжительность курса, количество лекций, тестов, результат теста и тип восприятия. На странице также пользователь может увидеть описание курса, индивидуальную учебную программу и участников курса в соответствии с рисунком 1.

Выводы

В данном диссертационном исследовании проводится разработка интеллектуальной информационно-обучающей системы с использованием моделей и методов семантического анализа. Теоретические исследования, проведенные в рамках диссертации, позволили получить следующие результаты, имеющие практическое и научное значение:

1. Анализ и обзор существующих систем дистанционного обучения и адаптивных интеллектуальных систем. Сформулированы требования к разработанной системе.

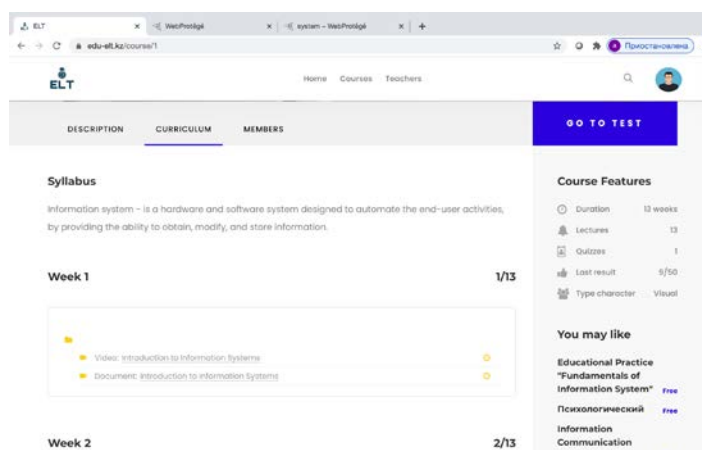


Рисунок 1 – Учебная программа курса

2. Используются научные методы для формирования контента, предназначенного для любого студента, который учитывает особенности пользователя.

3. Разработана методология построения образовательного ресурса на основе семантического анализа.

4. Выдвинута гипотеза о семантических связях между зависимыми от содержания и доказана в ходе исследования.

5. Предложена последовательность и управление количеством образовательных элементов интеллектуальной системы дистанционного обучения в области информационных технологий. Все компоненты этой системы подробно описаны. Математическая модель разработана с использованием семантического анализа данных.

6. Результат экспериментального исследования разработанной системы, демонстрирующий повышение качества образования и эффективности за счет семантического представления ресурсов, также создает условия для цифровизации образовательного процесса.

Достоверность теоретических результатов исследования подтверждается научными данными, экспериментальными исследованиями, а также промышленными испытаниями в реальных условиях.

Проведен расчет степени усвоения учебного материала. Оказалось, что в экспериментальной группе степень усвоения увеличилась на 0,62 балла, а в контрольной группе снизилась на 0,13 балла. Эффективность обучения с использованием системы в 1,15 раза выше. Значимость различий между контрольной и экспериментальной группами оценивали непараметрическим методом статистического анализа с использованием критерия Манна-Уитни. Экономическая эффективность результатов выполненных работ подтверждается актом внедрения.

В будущих работах можно рассмотреть рекомендации по использованию и созданию нового контента для обучения и рассмотреть возможность извлечения знаний из других типов данных, таких как аудио и видео, используемых в курсах.