



УДК 004.432.4

Г. Т. Бекманова

ЕНУ, г. Астана

**НЕКОТОРЫЕ ПОДХОДЫ К ПРОБЛЕМАМ АВТОМАТИЧЕСКОГО СЛОВОИЗМЕНЕНИЯ
И МОРФОЛОГИЧЕСКОГО АНАЛИЗА В КАЗАХСКОМ ЯЗЫКЕ**

Выявление формальных структур естественного языка, формализация языка в целом, построение конструктивной теории и компьютерной модели языка являются актуальными направлениями компьютерной лингвистики на протяжении последних десятилетий.

Наличие модуля словообразования в системах обработки естественно-языковых текстов приспособливает их к работе с неопределенными словами, которые получены путем использования различных словообразовательных моделей [1], например, сложение основ, конвертация частей речи, с помощью аффиксов и т.д. Системы автоматического словоизменения открывают большие возможности для углубленного изучения лексического состава языка, повышают грамотность, способствуют более быстрой и результативной работе с текстами, которые содержат новые, неизученные слова.

Модуль морфологического анализа является необходимым компонентом интеллектуальных систем, выполняющих обработку текста на естественном языке. Морфологический компонент оказывается необходимым и в орфографических корректорах, и в системах оптического распознавания символов, и в системах машинного перевода, и в системах распознавания речи. Особенно важным морфоанализ оказывается для языков с богатой морфологией, к которым относится и казахский язык.

Мы будем обсуждать указанную в заголовке тему, ориентируясь, прежде всего, на проблемы распознавания устной казахской речи. Естественный язык – это результат многовековой деятельности огромного числа носителей данного языка. Его предложения принципиально отличаются от случайных комбинаций слов и от предложений формально построенных языков. Одной из основных его особенностей является избыточность, позволяющая понимать искажённую речь.

Для многих языков, таких, как английский, французский, китайский, русский и т.п. разработаны различные языковые модели, принципиально улучшающие распознавание устной речи.

В состав любой модели системы распознавания речи входит словарь, размер которого влияет на степень сложности, требования к процедурам обработки и точность системы. Одним системам для работы необходимо всего несколько слов (например, только числа), а другие работают с очень большими словарями (например, диктофонные машины). Четких градаций объемов словарей нет. Но обычно словари классифицируются так:

- 1) маленькие словари — это десятки слов;

- 2) средние — сотни слов;
- 3) большие словари — тысячи слов;
- 4) очень большие словари — это десятки тысяч слов.

Начальные формы словаря, построенного автором настоящей работы, были получены из казахско-русского электронного словаря путем отделения казахской и русской частей. Необходимая морфологическая информация по русской части получена из электронного грамматического словаря Зализняка [2] и перенесена на казахские слова.

Особенностью казахского языка является наличие большого числа словоформ для каждого слова. Большинство словоформ одного и того же слова отличается только окончанием (или несколькими окончаниями) – это позволяет строить словарь автоматически, генерируя словоформы при помощи добавления окончаний.

Модель словоизменения базируется на присоединении окончаний на основе некоторых правил и признаков. Анализируется последний слог начальной формы слова (последняя буква слова и мягкость или твердость последнего слога). Это определяет признак (рис. 1), на основе которого происходит добавление окончания.

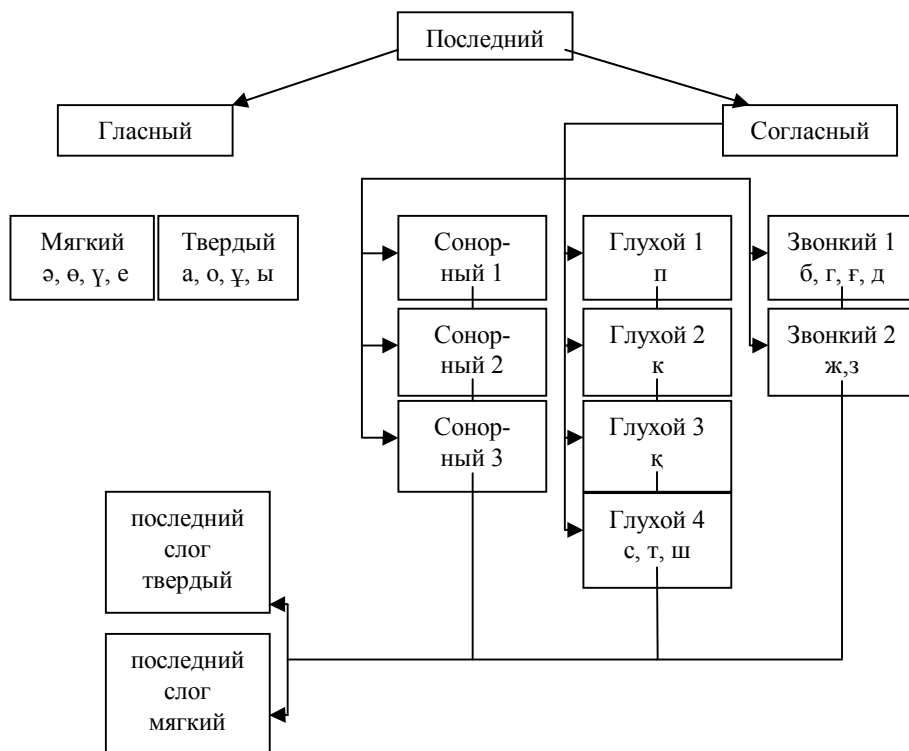


Рисунок 1- Признаки добавления окончаний существительных

В казахском языке строго определен порядок расположения окончаний. Например, для имен существительных к основе слова вначале добавляется окончание множественного числа затем притяжательное окончание (означает принадлежность предмета тому или иному лицу), далее следует падежное окончание и последним – окончание формы спря-

жения (добавляется только к одушевленным существительным) [3].

Таким образом, правило добавления окончаний можно представить в виде:

$$W = S \& K_g \& T_g \& S_g \& G_g \quad (1)$$

где W – словоформа; S – основа слова; K_g - окончание множественного числа; T_g - притяжательное окончание; S_g - падежное окончание; G_g - окончание формы спряжения.

Модель словоизменения можно представить в виде семантической нейронной сети, а сам словарь в виде синхронизированного линейного дерева.

Рассмотрим модель нейрона словарной статьи. Отдельная словарная статья представляет собой главное слово, несущее основную смысловую нагрузку статьи, и группу словоформ, получающуюся из главного слова путем словоизменения (спряжения или склонения).

Словарь начальных форм можно представить в виде модели нейрона. Начальную форму слова, из которой образуются все его словоформы путем склонения, спряжения и т.д., назовем леммой. Пусть лемма - это группа нейронов или один нейронный субавтомат.

При использовании синхронизированного линейного дерева для решения задач словоизменения/словообразования нейронная сеть будет выступать как переключательная цепь, коммутирующая возбуждение, переводящая субавтомат из одного состояния в другое. Переключение состояний субавтомата будет происходить при подаче на вход синхронизированного линейного дерева специальных команд. Эти команды будут распознаны синхронизированным линейным деревом и преобразованы в градиентное значение на выходе соответствующих им нейронов-эффекторов, что вызовет возбуждение или торможение нейронов, соответствующих состояниям леммы.

Внутренняя структура леммы в синхронизированном линейном дереве будет выглядеть следующим образом.

Таблица 1

Обозначения во внутренней структуре леммы

Символ	Обозначение
#	пробел между словами
!	начало слова
	конец слова
\$	морфологическая информация для образования словоформ

Рассмотрим пример для слова «окушы» (лемма) - «ученик» (в казахском языке одушевленные существительные изменяются по лицам с помощью личных окончаний). Рецептор возбуждается на символ начала слова «!». Далее переходит в состояние «о». При подаче символа «о», далее последовательно «!ок», «!оку», «!окуш», «!окушы», «!окушы|».

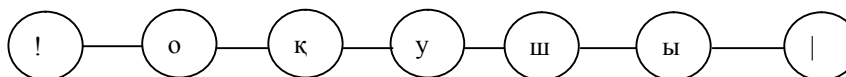


Рисунок 2 - Представление леммы в виде синхронизированного линейного дерева

Рассмотрим признаки лемм. Символом «*» обозначим первый спецсимвол признака леммы. Начало словоформы и ее признака обозначаем различными спецсимволами для уменьшения размера дерева поиска, это может увеличить скорость работы последовательной вычислительной системы. Однако, для решения задачи словоизменения на параллельной вычислительной системе было бы достаточно ограничиться различием спецсимволов «\$» и «|». На рис. 2 приведен пример структуры связей леммы, определяющей следующие признаки: зат есім (имя существительное) – «*зе\$», жанды (одушевленное) – «*жа\$», последний гласный символ «ы». В зависимости от этих семантических признаков выбирается траектория словоизменения. В первую очередь, имеет значение признак части речи, в зависимости от него мы словоизменяем слово в соответствии с правилами. Далее одушевленность – это признак, присущий исключительно существительным. Дело в том, что жіктік жалғау или окончание формы спряжения добавляется только к одушевленным существительным, т.е. можно сказать «Мен оқушымын» - «Я ученик», но нельзя сказать «Мен күрекпін» - «Я лопата». Семантические признаки, т.е. признаки которые выделяются у слов на основании их смысла, есть практически у каждой части речи. Поэтому их хранение в части морфологической информации необходимость. В таком случае можно представить лемму и его морфологическую информацию в виде синхронизированного линейного дерева, при подаче на вход которого специальных команд будет происходить переключение состояний субавтомата. Эти команды будут распознаны синхронизированным линейным деревом и преобразованы в градиентное значение на выходе соответствующих им нейронов-эффекторов, что вызовет возбуждение или торможение нейронов, соответствующих состояниям леммы. В случае возбуждения этих нейронов далее по правилам словоизменения будет производиться образование новых словоформ. Последний символ «ы» определяет тип окончания для существительного.

Аналогично можно представить модель морфологического анализа. Разница лишь в том, что в первом случае мы храним словарь начальных форм слов, а во втором - словарь словоформ.

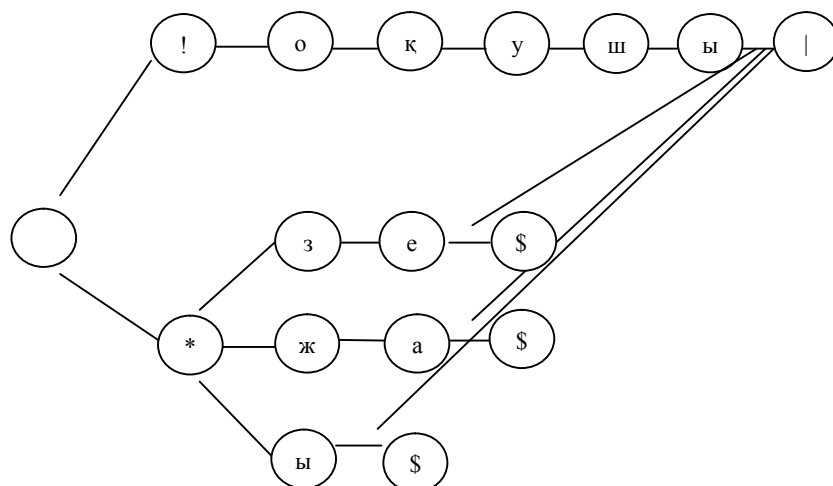


Рисунок 3 - Представление леммы и его морфологической информации в виде синхронизированного линейного дерева

Внутренняя структура словоформы в синхронизированном линейном дереве будет выглядеть следующим образом.

Таблица 2

Обозначения во внутренней структуре словоформы

Символ	Обозначение
#	пробел между словами
!	начало слова
	конец слова
\$	признак словоформы (падеж, число, спряжение и т. д.)

Рассмотрим пример для слова «бала» (лемма) - «ребенок» и двух его словоформ «балам» - «мой ребенок», «балаң» - «твой ребенок» (в казахском языке одушевленные существительные изменяются по лицам с помощью личных окончаний). Рецептор возбуждается на символ начала слова «!». Далее переходит в состояние «б». При подаче символа «б», далее последовательно «!ба», «!бал», «!бала» , и затем одновременно два субсостояния «!балам|» и «!балаң|» (рис. 4).

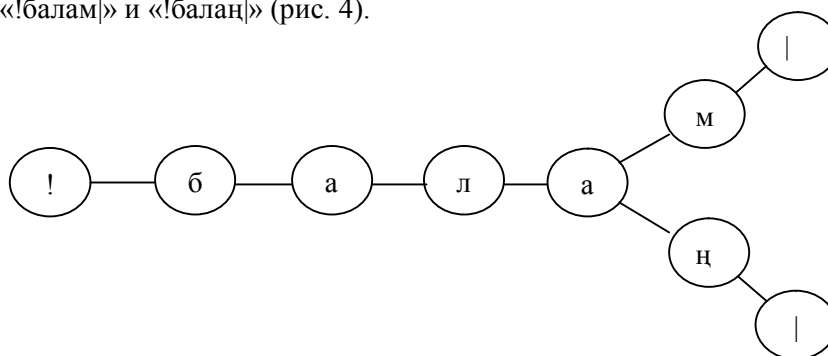


Рисунок 4 - Представление словоформ в виде синхронизированного линейного дерева

Рассмотрим признаки словоформ. Символом «*» обозначим первый спецсимвол признака словоформы, выступающий в той же роли, что и символ «!» для словоформ. Начало словоформы и ее признака обозначаем различными спецсимволами для уменьшения размера дерева поиска, это может увеличить скорость работы последовательной вычислительной системы. Однако для решения задачи словоизменения на параллельной вычислительной системе, было бы достаточно ограничиться различием спецсимволов «\$» и «|». На рис. 5 приведен пример структуры связей леммы, определяющей следующие признаки: зат есім (имя существительное) –«*зе\$», жіктік жалғау (личное окончание), бірінші бет (1 б на рис. 5, переводится как первое лицо), жіктік жалғау (личное окончание), екінші бет (2 б на рис. 5, переводится как второе лицо). При подаче на лемму слова «.балам_» она переходит в возбужденные субсостояния: «!балам|», «*зе\$», «*ж1бже\$», а

при подаче слова «балац» - в возбужденные субсостояния: «!балац|», «*зе\$», «*ж2бже\$» [4].

Данный метод морфологического анализа, когда на диске хранится весь словарь, является декларативным. Существуют также процедурный и комбинированный методы морфологического анализа [5]. Процедурный морфологический анализатор состоит из следующих этапов: выделение в текущей словоформе основы, ее идентификация, приписывание словоформе соответствующего перечня морфологической информации. Комбинированный метод сочетает в себе декларативный и процедурный и чаще всего используется на практике.

Модули словоизменения и морфологического анализа являются неотъемлемой частью современных интеллектуальных информационно-поисковых, диалоговых и обучающих систем, средств машинного перевода, модулей проверки правописания, распознавания речи, в которых используется анализ естественно-языковых текстов.

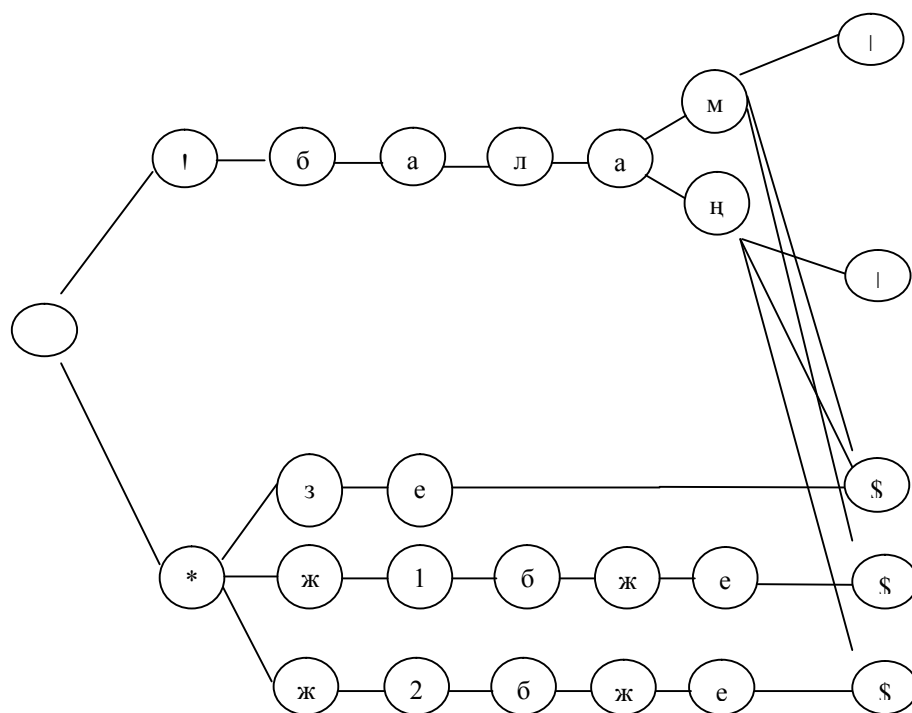


Рисунок 5 - Представление словоформ и их морфологической информации в виде синхронизированного линейного дерева

Список литературы

1. Вороной С.М. Повышение эффективности интеллектуального поиска в полнотекстовых базах данных на основе автоматического аннотирования документов /С.М. Вороной, А.А. Егошина; Под ред. Т.А. Таран // VII Международная конференция «Интеллектуальный анализ информации ИАИ-2007».: /Рос. ассоц. искусств. интеллекта и др. -Киев: Просвіта, 2007. - 392 с.: илл.
2. [Зализняк](#) А.А. Грамматический словарь русского языка. -М.: Русский язык, 1987.-880 с.
3. Туймебаев Ж.К. Казахский язык. Грамматический справочник. - Алматы: Гылым, 2004.-

108 с.

4. Шуклин Д.Е. Структура семантической нейронной сети, реализующей морфологический и синтаксический разбор текста.
5. Ножов И.М. Прикладной морфологический анализ без словаря. //КИИ-2000. Труды конференции – М.: Физматлит, 2000. – Т.1. –С. 424-429.

Получено 26.10.09

УДК 574.721

Н.А.Нурбаева

ВКГТУ, Усть-Каменогорск

ТЕХНОЛОГИЯ ДИСТАНЦИОННОГО ОБУЧЕНИЯ В СИСТЕМЕ ОБРАЗОВАНИЯ И ВОСПИТАНИЯ СТУДЕНТОВ

В условиях быстроменяющегося мира невозможно получить образование на всю жизнь, поскольку только путем обучения на протяжении всей жизни возможно сохранение и повышение компетентности и профессионализма специалиста. Конкурентоспособный специалист, стремящийся к профессиональному росту, ищет пути получения новых знаний без отрыва от основного вида деятельности. Вуз, учитывая повышение требований к специалисту, ищет новые формы и методы обучения, разрабатывает и внедряет новые технологии в образовательную деятельность. Одной из новых форм организации учебного процесса является обучение посредством дистанционных образовательных технологий (ДОТ). ДОТ – это технологии, которые позволяют с одной стороны удовлетворить образовательные потребности обучаемых, с другой - привлечь большее количество студентов в вуз, что позволяет достигнуть взаимного равновесия между потребителем и поставщиком образовательных услуг.

Современное состояние науки и образования, обусловленное экономическими и социальными факторами характеризуется повышением требований и качеству подготовки специалистов и определяет необходимость постоянного поиска новых методов и средств повышения эффективности образовательного процесса. Одним из таких методов является дистанционное обучение (ДО). Технологии ДО обеспечивают адаптацию процесса обучения к индивидуальным характеристикам обучаемых, освобождают преподавателей от ряда трудоемких и часто повторяющихся операций по представлению учебной информации и контролю знаний и отличают накопление учебно-методического опыта. Анализ ряда научно-педагогических и технических трудов и результаты исследований позволяют сделать вывод об эффективности технологий ДО в образовательной деятельности. С целью обоснования использования ДОТ в системе воспитания приведем результаты анализа основных преимуществ и недостатков этих технологий.

Преимущества технологий ДО: гибкость (возможность заниматься в удобное для себя

время, в удобном месте, темпе); модульность и вариативность (возможность из набора независимых учебных курсов (модулей) формировать программу обучения, отвечающую индивидуальным или групповым потребностям); параллельность (обучение параллельное с профессиональной деятельностью); возможность постоянного повышения своего образовательного уровня. Многообразие источников информации (электронные библиотеки, банки данных, базы знаний и т.д.), технологичность (использование в образовательном процессе новейших достижений, современных информационных технологий), модифицируемость (относительная легкость обновления содержания учебного материала, возможности архивации старого материала), доступность (любой учебный материал остается у обучаемого фиксированным в виде компьютерных лекций и может быть в любой момент востребован), повышение учебной мотивации (стимулирование самостоятельности в обучении, умение критически мыслить, самодисциплина и ответственность, настойчивость в достижении цели), экономичность (эффективность использования учебных площадей, технических средств; концентрированное, унифицированное представление учебной информации и мультидоступ к ней; снижающие затраты на подготовку специалистов).

Недостатки технологий ДО: существует ряд практических навыков, которые можно получить только при выполнении реальных практических работ; успешность обучения частично зависит от навыков управления компьютером, перемещения в Интернете и от способности справляться с техническими трудностями, дефицитом доверия к электронным средствам общения и обучения; слушатели хотят видеть преподавателя и общаться с ним «вживую»; недостаточная интерактивность, сложность восприятия больших объемов информации с экрана; ослабление контроля со стороны преподавателя; снижается разнообразие форм учебного процесса и плюрализм мнений; создание дистанционных курсов дорого и занимает достаточно много времени; программно-методические требования, предъявляемые к учебно-методическим пособиям, ограничивают возможность авторов наиболее полно представить изучаемый материал, который в отсутствие преподавателя полностью берёт на себя функции управления образовательным процессом; достаточно высокая стоимость высокопроизводительных программно-технических и телекоммуникационных средств.

Названные недостатки носят достаточно объективный характер, хотя некоторые из них могут быть устранены различными способами: техническими, организационными, методическими, дидактическими, функциональными. Например, недостаточные навыки работы на компьютере компенсируются наличием максимально упрощенного и понятного пользовательского интерфейса. Недостаточная интерактивность восполняется за счет использования определенных дидактических приемов и интеллектуальных технологий моделирования знаний и деятельности по организации консультаций в очной и дистанционной форме. Ослабление контроля со стороны преподавателя компенсируют сеансы дистанционного контроля, проводимые по жесткому графику и предусматривающие представление обучаемыми отчетных материалов по каждому пройденному разделу.

Приведенные примеры свидетельствуют о целесообразности использования техноло-

гий ДО в комплексе с традиционными учебно-методическими средствами. При этом понятно, что ДО не является исключительной альтернативой традиционного обучения, а наоборот, должно органично встраиваться в существующую образовательную систему.

Обучаться с применением ДО могут студенты любой формы обучения (очной, заочной), с любым базовым образованием (школа, колледж, высшее образование), по любой специальности вуза, оформившие заявление в деканате факультета, а также студенты, обучающиеся в рамках международного сотрудничества (далее зарубежье) и в центрах ДОТ.

Учебный процесс организуется на основе рабочего учебного плана ДОТ, составленного в соответствии с требованиями государственного общеобязательного стандарта, утвержденного руководителем организации образования. Организация образования использует в образовательном процессе те виды проведения учебных занятий, которые соответствуют выбранным технологиям ДО.

Индивидуальный учебный план составляется обучающимся, согласовывается и утверждается руководителем ДОТ. Учебный год в организации образования начинается в соответствии со сроком, установленным Министерством образования и науки Республики Казахстан и состоит из академических периодов, количество и продолжительность которых устанавливается организацией образования самостоятельно. Обучающийся по ДОТ должен полностью освоить содержание профессиональных учебных программ в соответствии с государственными общеобязательными стандартами образования.

Организация образования, реализующая учебный процесс по ДОТ, обеспечивает научную и методическую помощь обучающимся с учетом их возможностей, через взаимодействие участников посредством информационно-телекоммуникационных средств.

Самостоятельная работа обучающегося под руководством преподавателя-тьютора включает интерактивные консультации по всем учебным материалам дисциплины (посредством чата, форума) и самостоятельное изучение обучающимися теоретического материала. Консультации при реализации образовательных учебных программ с применением ДОТ могут проводиться в различных формах: очные индивидуальные (по инициативе обучающегося); дистанционные индивидуальные (рецензии на контрольные и аттестационные работы, по телефону, через e-mail, чат, форум); дистанционные групповые (через чат, рубрику FAQ - часто задаваемых вопросов на Web-сайте, в виде телеконференций). При ДО учебные занятия проводятся в форме видеолекций, мультимедиа-лекций; самостоятельной работы обучающихся по электронным учебным изданиям; компьютерных тестирующих систем; консультации через форумы, чаты и электронную почту.

Организация образования, реализующая ДОТ, должна осуществлять текущий контроль успеваемости и промежуточную аттестацию обучающихся в информационной системе, защищенной против подлогов, фальсификаций путем применения электронных методов ограничения доступа, методов независимой оценки, публичной защиты работ в территориальных аттестационных комиссиях. Организация образования обеспечивает обучающихся возможностью прохождения всех видов профессиональной практики, предусмотренных государственными общеобязательными стандартами образования.

Получено 30.10.09
