

**ГОУ ВПО РОССИЙСКО-АРМЯНСКИЙ (СЛАВЯНСКИЙ)
УНИВЕРСИТЕТ**

Составлен в соответствии с
государственными требованиями к
минимуму содержания и уровню
подготовки выпускников по
направлению _____ и
Положением «Об УМКД РАУ».

УТВЕРЖДАЮ:
Директор института
Саркисян Г.З.
“27” декабря 2014 г.

Институт: Институт Гуманитарных наук

Кафедра: Мировой литературы и культуры

Автор(ы): К. ф. н., Татевосян Анаит Александровна

УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС

Дисциплина: **М2.Б.1 Компьютерные технологии в филологии**

Магистерская программа: 032700.68

Направление: Филология

ЕРЕВАН

1. Аннотация

1.1. Выписка из ФГОС ВПО РФ по минимальным требованиям к дисциплине

В результате изучения дисциплины студент должен:

знать: современные информационные технологии, применяемые при сборе, хранении, обработке анализе и передаче филологической информации;

уметь: самостоятельно использовать современные информационные технологии для решения научно-исследовательских и производственно-технологических задач профессиональной деятельности;

владеть: коммуникативными стратегиями и тактиками, риторическими, стилистическими и языковыми нормами и приемами, принятыми в разных сферах коммуникации, адекватно использовать их при решении профессиональных задач.

1.2. Взаимосвязь дисциплины с другими дисциплинами учебного плана специальности (направления)

Курс «Компьютерные технологии в филологии» взаимосвязан с дисциплинами:

- «Математика и информатика» (Б2.Б 3);

1.3. Требования к исходным уровням знаний, умений и навыков студентов для прохождения дисциплины (что должен знать, уметь и владеть студент для прохождения данной дисциплины)

Для прохождения курса «Компьютерные технологии в филологии» студент должен освоить программу бакалавриата по специальности «филология», поскольку приобретенные навыки решения филологических задач с использованием компьютерных технологий должны применяться студентом в рамках профессиональной филологической деятельности.

1.4. Предварительное условие для прохождения (дисциплина(ы), изучение которых является необходимой базой для освоения данной дисциплины)

Предварительным условием для прохождения курса «Компьютерные технологии в филологии» является освоение курса информатики, базовые навыки работы за компьютером, владение базовыми филологическими знаниями и навыками в рамках программы бакалавриата.

2. Содержание

2.1. Цели и задачи дисциплины

Важнейшей особенностью современной образовательной системы является применение новых информационных систем и технологий. Проекты по их созданию являются приоритетным и наиболее перспективным направлением в развитии гуманитарной сферы в целом и филологической в частности. Создаются обучающие программы, проводится компьютерное тестирование, разрабатываются программы для дистанционного обучения, реализуются различные методы анализа текстов. Новые технологии открывают более широкие в сравнении с традиционным обучением возможности в преподавании и изучении филологических дисциплин, в особенности в преподавании языков. Для эффективного использования предлагаемых новыми технологиями возможностей необходимо полноценное представление об имеющихся в Интернете ресурсах и базах данных предназначенных для филологов и программном обеспечении, как облегчающем профессиональную деятельность филологов, так и предназначенном для обучения.

Цель дисциплины – создание навыков работы с предназначенными для филологов сетевыми ресурсами и программным обеспечением для дальнейшего использования их при решении профессиональных задач и в преподавании.

Задачи дисциплины:

- ознакомление студентов с предназначенными для филологов ресурсами: электронными библиотеками, информационными проектами, базами данных;
- ознакомление студентов с методами автоматического анализа текста на морфологическом, синтаксическом и семантическом уровнях;
- изучение соответствующих программных средств;
- изучение форм использования компьютерных технологий в филологическом образовании.

2.2. Требования к уровню освоения содержания дисциплины (какие компетенции (знания, умения и навыки) должны быть сформированы у студента ПОСЛЕ прохождения данной дисциплины)

После прохождения курса «Компьютерные технологии в филологии» студент должен:

- ориентироваться в предназначенных для филологов сетевых ресурса,
- уметь анализировать текст с помощью соответствующего программного обеспечения,

- уметь применять компьютерные технологии в профессиональной деятельности.

2.3. Трудоемкость дисциплины и виды учебной работы (в академических часах и кредитах)

2.3.1. Объем дисциплины и виды учебной работы

Виды учебной работы	Всего, в акад. часах
1. Общая трудоемкость изучения дисциплины по семестрам, в т. ч.:	72
1.1. Аудиторные занятия, в т. ч.:	36
1.1.1. Лекции	
1.1.2. Практические занятия, в т. ч.	36
1.1.2.1. Обсуждение прикладных проектов	
1.1.2.2. Кейсы	
1.1.2.3. Деловые игры, тренинги	
1.1.2.4. Контрольные работы	2
1.1.2.5. Другое (указать)	
1.1.3. Семинары	
1.1.4. Лабораторные работы	
1.1.5. Другие виды (указать)	
1.2. Самостоятельная работа, в т. ч.:	36
1.2.1. Подготовка к экзаменам	
1.2.2. Другие виды самостоятельной работы, в т.ч. (указать)	
1.2.2.1. Письменные домашние задания	
1.2.2.2. Курсовые работы	
1.2.2.3. Эссе и рефераты	
1.2.2.4. Другое (указать)	
1.3. Консультации	
1.4. Другие методы и формы занятий	
Итоговый контроль (экзамен, зачет, диф. зачет - указать)	зачет

2.3.2. Распределение объема дисциплины по темам и видам учебной работы

Разделы и темы дисциплины	Всего (ак. часов)	Лекции (ак. часов)	Практ. занятия (ак. часов)	Семина-ры (ак. часов)	Лабор. (ак. часов)	Другие виды занятий (ак. часов)
1	2=3+4+5+6+7	3	4	5	6	7
Раздел 1. Базы данных и Введение.	8					
Тема 1.1. Электронные библиотеки художественной литературы и научных изданий			2			
Тема 1.2. Филологические информационные проекты по созданию электронных словарей			2			
Тема 1.2. Ресурсы, предназначенные для атрибуции литературных произведений.			2			
Раздел 2. Компьютерные технологии обработки данных	12					
Тема 2.1. Программы анализа и лингвистической обработки			4			
Тема 2.2. Программы, осуществляющие машинный перевод.			4			
Тема 2.3. Генераторы текстов.			4			
Тема 2.4. Программы, обеспечивающие стилистический анализ.			2			
Раздел 3. Использование компьютерных программ и Интернета в филологическом образовании.	14					
Тема 3.1. Системы дистанционного обучения.			4			
Тема 3.2. Гипертекстовые системы.			2			
Тема 3.3. Типы компьютерных обучающих программ.			4			
Тема 3.4. Типы электронных учебников.			4			
Контрольная						2
ИТОГО	36		36			

2.3.3 Содержание разделов и тем дисциплины

Содержание курса с гиперссылками на изучаемые ресурсы выложено на моей страничке в интернете по адресу <http://www.diary.ru/~flamto> для того, чтобы у магистрантов был к нему и внеаудиторный доступ.

Раздел 1. Базы данных и филологические ресурсы интернета.

Введение

Работа с базами данных и филологическими ресурсами в Интернете. Необходимость использования компьютерного обеспечения для обработки и презентации материалов исследований и педагогической деятельности. Использование цифровых продуктов в образовании.

Тема 1.1. Электронные библиотеки художественной литературы и научных изданий.

Каталог электронных библиотек **Library.Ru**: <http://www.library.ru/2/catalogs/elibs>;
Библиотека Максима Мошкова, крупнейшая бесплатная электронная библиотека рунета: <http://lib.ru>; электронная библиотека **Im Werden**: <http://imwerden.de/cat/modules.php?name=books>; **Классика.ру**, библиотека классической литературы: <http://www.klassika.ru>; Древнерусская литература: <http://old-rus.narod.ru>;
Вавилон, библиотека современной русской литературы <http://vavilon.ru>; электронные научные издания различных разделов **Фундаментальной электронной библиотеки (ФЭБ)** «Пушкин», «Лермонтов», «Толстой», «Русская литература и фольклор» feb-web.ru/; конкордансы всех произведений Ф. М. Достоевского, В.И. Даля www.philolog.ru; **РГБ**, электронная библиотека Российской государственной библиотеки: <http://www.rsl.ru/ru/s2/s101/>; **Библиотека Гумер**, коллекция книг по социальным и гуманитарным и наукам <http://www.gumer.info/>; **E-Lingvo.net**, проект посвященный языкознанию, литературе, литературоведению <http://e-lingvo.net/files>; **Электронная библиотека учебников**, коллекция учебной литературы по социально-экономическим и гуманитарным наукам: <http://studentam.net>.

Тема 1.2. Филологические информационные проекты по созданию электронных словарей различного типа, научно-справочных и исследовательских баз данных.

Научно-образовательный портал «**Лингвистика в России: ресурсы для исследователей**»: uisrussia.msu.ru/linguist/index.jsp; проект РАН и Института языка им. В.В. Виноградова: «Этимология и история русского языка»: etymolog.ruslang.ru; проект филологического факультета МГУ «**Русская фонетика**»: fonetica.philol.msu.ru; филологический портал www.philology.ru; **Рутения**, единый информационный ресурс, ориентированный на исследователей-русистов www.ruthenia.ru; словарь литературоведческих терминов slovar.by.ru; справочно-информационный портал **ГРАМОТА.РУ**: gramota.ru; все о языках, лингвистике, переводе linguistic.ru; информационные ресурсы для филологов www.lib.csu.ru/polssilki/philology.shtml; культура письменной речи www.grammar.ru; сайт студентов-филологов www.durov.com/.

Тема 1.3. Ресурсы, предназначенные для атрибуции литературных произведений.

Информационная система «**Статистические методы анализа литературного текста**»: smalt.karelia.ru/index.html; **Лингвоанализатор**: www.rusf.ru/books/analysis/history.htm.

Раздел 2. Компьютерные технологии обработки данных.

Тема 2.1. Программы анализа и лингвистической обработки текстов.

Russian Morphological Dictionary - программа для синтаксического и морфологического анализа русскоязычных текстов. **Система StarLing**, морфологический анализатор С.А.Старостина для слов русского/английского языков на основе словарей Зализняка (рус.яз.) и Мюллера (англ.яз.). Позволяет получить для вводимого слова базовую форму и морфологическую информацию. **Программа mystem** - производит морфологический анализ текста на русском языке. Для слов, отсутствующих в словаре, порождаются гипотезы. Программные продукты **АОТ (Автоматическая обработка текста)**:

модуль графематического анализа текста; компоненты морфологического анализа для русск., нем. и англ. яз.; модуль автоматического уничтожения омонимии; система синтаксического и фрагментационного анализа для русск., и нем. яз.; система первичного семантического анализа. **TextAnalyst** – выделяет из обрабатываемых текстов основные понятия, ранжирует их по значимости, устанавливает связи между ними и строит дерево терминов. Позволяет осуществлять навигацию по тексту при помощи дерева терминов. Имеет функции ассоциативного поиска. **Синтаксический анализатор естественного текста на русском языке** – программа, обрабатывающая предложения с большим количеством слов (100 и более). **«Морфер»** – программный модуль, выполняющий склонение ФИО, наименований должностей, профессий, организаций, продуктов и почти всего, что склоняется в русском языке. **Склонятель** – программа для склонения наименований на русском языке.

Тема 2.2. Программы, осуществляющие машинный перевод.

Prompt – онлайн переводчик, интегрирующий возможности технологий Translation Memory и машинного перевода, поддерживает русский, английский, немецкий, французский и испанский языки, др. программы-переводчики. **Мультитран** – система автоматического перевода, поддерживающая русский, английский, немецкий, французский и испанский языки, предоставляющая возможности алфавитного, морфологического и фразового поиска. Online-переводчики текстов и web-страниц на русский язык. Переводчик **Google**, система статистического машинного перевода.

Тема 2.3. Генераторы текстов.

Тест Тьюринга. Эмуляторы искусственного интеллекта. **Chat Master, A. L. I. C. E. Artificial Linguistic Internet Computer Entity** и другие русскоязычные самообучающиеся программы, ведущие контекстно-зависимый диалог. **Scott Pakin's automatic complaint-letter generator** – генератор письменных жалоб Скотта Пейкин (генерирует тексты жалоб на заданную персону или организацию). **Russian Word Constructor (RWC)**, программа, конструирующая русские неологизмы на основе словаря с лексико-статистической информацией о языке.

Тема 2.4. Программы, обеспечивающие стилистический анализ.

Fresh Eye (Свежий Взгляд) – утилита для выбраковки расположенных близко в тексте фонетически и морфологически сходных слов. **Худломер** – программа, осуществляющая автоматическую классификацию функционального стиля текста на основе спектров длин слов.

Раздел 3. Использование компьютерных программ и Интернета в филологическом образовании.

Тема 3.1. Системы дистанционного обучения.

Кейс-технология. Локальные и сетевые СДО. Интегрированные обучающие среды **WebCT** и **LearningSpace**. Неинтегрированные программные пакеты для создания упражнений, проведения тестирования, организации работы в группах и поддержки интерактивного обучения.

Тема 3.2. Гипертекстовые системы.

World Wide Web – гипертекстовая система, основанная на клиент-серверной архитектуре и работающая на Интернет. **HyperWave** – система управления документами Web в больших информационных пространствах. **Microcosm** – открытая гипермедиа система для разработки онлайн-учебников, справочников и документации. **Storyspace** – система, поддерживающая процесс написания гипертекстовых произведений («писательская среда»). **WebThing** – объектно-ориентированная гипертекстовая система, спроектированная для совместной авторской работы. Линейное и структурное представление мультимедиа-информации

Тема 3.3. Типы компьютерных обучающих программ.

Тренировочные и контролирующие (предназначены для закрепления умений и навыков); наставнические (предлагают учащимся теоретический материал для изучения); имитационные и моделирующие (позволяют осуществлять компьютерный эксперимент); развивающие игры (предоставляют в распоряжение учащегося условную среду, набор возможностей и средства их реализации).

Тема 3.4. Типы электронных учебников.

Тест. Энциклопедия. Задачник. Креативная среда. Авторская среда. Невербальная среда.

2.4. Материально-техническое обеспечение дисциплины

Необходима компьютерная аудитория с интернетом.

2.5. Распределение весов по модулям и формам контроля

Формы контролей	Весы форм текущих контролей в результирующих оценках текущих контролей			Весы форм промежуточных контролей в оценках промежуточных контролей			Весы оценок промежуточных контролей и результирующих оценок текущих контролей в итоговых оценках промежуточных контролей			Весы итоговых оценок промежуточных контролей в результирующей оценке промежуточных контролей	Весы результирующей оценки промежуточных контролей и оценки итогового контроля в результирующей оценке итогового контроля	
	M1 ¹	M2	M3	M1	M2	M3	M1	M2	M3			
Вид учебной работы/контроля												
Контрольная работа						1						
Тест												
Устный опрос												
Курсовая работа												
Лабораторные работы												
Письменные домашние задания												
Реферат												
Эссе												

¹ Учебный Модуль

Другие формы (Указать)											
Другие формы (Указать)											
Весы результирующих оценок текущих контролей в итоговых оценках промежуточных контролей											
Весы оценок промежуточных контролей в итоговых оценках промежуточных контролей											
Вес итоговой оценки 1-го промежуточного контроля в результирующей оценке промежуточных контролей										1	
Вес итоговой оценки 2-го промежуточного контроля в результирующей оценке промежуточных контролей											
Вес итоговой оценки 3-го промежуточного контроля в результирующей оценке промежуточных контролей											
Вес результирующей оценки промежуточных контролей в результирующей оценке итогового контроля											0,5
Экзамен/зачет (оценка итогового контроля)											0,5 (Зачет)
	$\Sigma=1$	$\Sigma=1$	$\Sigma=1$	$\Sigma=1$	$\Sigma=1$	$\Sigma=1$	$\Sigma=1$	$\Sigma=1$	$\Sigma=1$	$\Sigma=1$	$\Sigma=1$

3. Теоретический блок

3.1. Материалы по теоретической части курса

3.1.1. Учебник:

1. Компьютерные технологии в филологии: http://ru.wikibooks.org/wiki/Компьютерные_технологии_в_филологии

3.1.2. Учебные пособия:

1. Потапова Р.К. - Новые информационные технологии и лингвистика. Любое издание.
2. Шемякин Ю.И. Начала компьютерной лингвистики: Учеб. пособие. М.: Изд-во МГОУ, А/О "Росвузнаука", 1992.
3. Кедрова Г.Е. Методы оптимизации компьютерной обучающей среды по лингвистике для систем дистанционного обучения в интернете, <http://www.philol.msu.ru/~kedr/kedr-ulj.htm>

4. Эпштейн В.Л. Введение в гипертекст и гипертекстовые системы.
<http://www.ipu.ru/publ/epstn.htm>.

3.1.3. Конспект лекции:

Хотя еще в 1936 г. философ Альфред Айер предложил способ отличения человека от машины: «Единственным основанием, на котором я могу утверждать, что объект, который кажется разумным, на самом деле не разумное существо, а просто глупая машина, является то, что он не может пройти один из эмпирических тестов, согласно которым определяется наличие или отсутствие сознания», тест получил имя предложившего его в 1950 г. Алана Тьюринга. Тьюринг писал: «Нетрудно разработать машину, которая будет неплохо играть в шахматы. Теперь возьмем трех человек — субъектов эксперимента. А, В и С. Пусть А и С неважно играют в шахматы, а В — оператор машины. <...> Используются две комнаты, а также некоторый механизм для передачи сообщений о ходах. Участник С играет или с А, или с машиной. Участник С может затрудниться ответить с кем он играет». Сам тест создан по аналогии с игрой «Imitation game», в которой мужчину и женщину разводят по разным комнатам, куда им передают записанные вопросы и получают на бумажках ответы, а остальные участники должны отгадать, кто из них кто, при том, что оба участника стараются их запутать. В общем виде «тест Тьюринга» формулируется так: «Человек взаимодействует с одним компьютером и одним человеком. На основании ответов на вопросы он должен определить, с кем он разговаривает: с человеком или компьютерной программой. Задача компьютерной программы — ввести человека в заблуждение, заставив сделать неверный выбор».

Человек, принимающий решение, получает ответы обоих своих собеседников отпечатанными, через посредника (человека или компьютер), чтобы он не мог ориентироваться на продолжительность времени, которое требуется собеседникам, чтобы ответить на вопросы (первоначально это условие было введено потому, что компьютер давал ответ намного медленнее человека, сейчас оно необходимо потому, что медленнее реагирует человек). Если по итогам беседы он не может с уверенностью сказать, кто из его собеседников человек, а кто компьютер — считается, что тест пройден. До сих пор ни одна программа-собеседник тест Тьюринга не прошла.

С точки зрения Б. Витби история утверждения теста Тьюринга определяется следующими поворотными пунктами: публикацией статьи «Вычислительные машины и разум» (1950 г.), где эта идея было впервые сформулирована, сообщением о создании Дж.

Вейценбаумом программы «Элиза» (1966 г.), созданием К. Колби программы PARRY (1972 г.), и Коллоквиумом Тьюринга (1990 г.).

«Элиза» определяет во введенном в нее сообщении ключевые слова, после чего сообщение преобразуется по определенному алгоритму и возвращается автору. Если программа не может определить ключевое слово, то либо использует один из введенных в нее общих ответов, либо повторяет одно из предыдущих сообщений. Первоначально программа имитировала поведение психотерапевта, за счет чего некоторые из испытуемых верили, что она человек, однако они не были предупреждены о том, что могут иметь дело с машиной, поэтому зафиксированные случаи не считаются прохождением программой теста. Программа «PARRY» имитировала поведение параноидального шизофреника, наряду с реальными шизофрениками проходя тест, проводимый командой психиатров. Затем другая команда психиатров перепроверяла результат по стенограммам. В 48 % психиатры верно определили, кто из испытуемых – машина. Однако у тех, кто принимал решение не было возможности задать машине вопрос, они работали со стенограммами уже проведенной беседы, поэтому в этом случае тоже не считается, что программа прошла тест.

В 1990 г., после коллоквиум Тьюринга, на котором был обсужден тест и его перспективы, была учреждена ежегодно присуждаемая премия Лёбнера, которая вручается лучшей, «продемонстрировавшей наиболее человеческое поведение» программе по итогам теста. Серебряная и золотая медали этой премии так ни разу и не были присуждены, однако бронзовая присуждается каждый год. Так, программа (A.L.I.C.E. – Artificial Linguistic Internet Computer Entity, «Искусственное лингвистическое интернет-компьютерное существо») получала ее в 2000, 2001 и 2004, Jabberwacky – в 2005 и 2006. Последней в 2010 г. приз получила программа Suzette, разработанная Брюсом Уилкоксом.

Тест Тьюринга подвергался критике на различных основаниях. В частности, для прохождения теста машина должна быть не разумной, а антропоморфной: она должна не демонстрировать разумность, а имитировать поведение и реакции человека. Между тем, человек бывает нелогичным, агрессивным, эмоциональным – и машина должна имитировать в том числе и неадекватное поведение человека, чтобы пройти тест. То есть для прохождения теста компьютер надо научить не хорошо думать, а хорошо обманывать. Когда же он демонстрирует разумность – решает сложные задачи, которые не мог быть решить человек, тест он не проходит. Кроме того, те сферы, в которых требуются разумные компьютеры, вовсе не требуют от компьютера имитации человеческого поведения.

На сегодняшний день активно используется также т.н. обратный тест Тьюринга или САРТСНА. Это требование ввести некий набор букв и цифр с картинки, чтобы, при авторизации, доказать, что вы являетесь человеком, а не программой.

На сегодняшний день существует большое количество программ-эмуляторов искусственного интеллекта. Не все из них действительно анализируют введенный текст. Так, программа "Элочка", имитирующая диалог с героиней романа «12 стульев», словарный запас которой включал 30 слов, выдает этот набор слов в произвольном порядке, адекватно выполняя, при этом, поставленную перед ней задачу.

В целом же программа анализирует введенное сообщение, разбивает на простые предложения, определяет ключевые слова и сочетания, сравнивает введенный текст с эталонами из базы, подбирая максимально близкий образец, определяет тематику сообщения, подбирает синонимы, фильтрует иррелевантные элементы и затем, исходя из полученного результата выдает либо стандартный ответ из базы, если фраза оказывается близкой к какому-либо образцу, либо выбирает ответ, ориентируясь на ключевые слова, либо, если ключевые слова выделить не может, ориентируется на тематику разговора, учитывая предыдущие сообщения. Часть эмуляторов искусственного интеллекта обучаемы. Они «запоминают» использованные в беседе с ними сообщения или могут работать в режиме «обучающего диалога», когда пользователь вводит вопрос, а затем желаемый ответ, который он в дальнейшем хотел бы получать на этот вопрос от программы. Адекватность ответов зависит от уровня анализа исходного сообщения, от объема базы и ее функциональности (база, из которой программа выбирает фразы достаточно произвольно, ничуть не улучшает качество диалога), от наличия некоторого количества ответов на стандартные, часто встречающиеся вопросы и от способности программы «понимать» слова, написанные с распространенными ошибками.

4. Практический блок

4.1. Задания по курсу:

1. Подготовка текстов для работы с программами анализа и лингвистической обработки.
2. Подготовка текстов для работы с программами стилистического анализа.

Выполнение этих двух заданий не оценивается.

5. Материалы по оценке и контролю знаний

5.1. Задания по курсу:

1. Подготовка текстов, которые адекватно, без ошибок переводятся на другой язык как стандартной программой-переводчиком, так и системой статистического машинного перевода. Для выполнения задания необходимо четко представлять себе используемые соответствующими программами принципы подстановки.

2. Подготовка условного дистанционного курса (на базе дипломной работы) состоящего из блоков, связанных системой гиперссылок.
3. Подготовка электронного теста из десяти вопросов (на базе созданного в рамках предыдущего задания дистанционного курса).

Средняя оценка за выполнение этих трех заданий является итоговой оценкой по курсу.

6. Методический блок

6.1. Методика преподавания, обоснование выбора данной методики

Освоение дисциплины предполагает следующие формы работы преподавателя:

– проведение практических занятий;

В рамках практических занятий изучаются ресурсы, усваиваются принципы работы программ, выполняются задания по курсу.

6.2. Методические рекомендации для студентов

Самостоятельная работа магистрантов предполагает просмотр предложенных ресурсов, выполнение заданий по курсу.

6.3. Методические указания по организации самостоятельной работы студентов при изучении конкретной дисциплины

В соответствии с учебным планом по курсу предусмотрен *промежуточный контроль* (осуществляется в конце учебного семестра в форме выполнения задания по переводу).

Итоговая оценка знаний студента выводится на основании следующих критериев: активность участия в практических занятиях, выполнение заданий.