

ՀՀ ԳԱԱ ԻՆՖՈՐՄԱՏԻԿԱՅԻ ԵՎ ԱՎՏՈՄԱՏԱՑՄԱՆ ՊՐՈԲԼԵՄՆԵՐԻ ԻՆՍՏԻՏՈՒՏ

Թոփչյան Տիգրան Ռուբենի

ԲԻԶՆԵՍ ՊՐՈՑԵՍՆԵՐԻ ՕՊՏԻՄԻԶԱՑԻԱՅԻ ԽՆԴԻՐՆԵՐԸ ԿՐԹԱԿԱՆ ՀԱՄԱԿԱՐԳԻ
ՕՐԻՆԱԿԻ ՎՐԱ

Ե.13.04 – «Հաշվողական մեքենաների, համալիրների, համակարգերի և ցանցերի մաթեմատիկական և ծրագրային ապահովում» մասնագիտությամբ տեխնիկական գիտությունների թեկնածուի գիտական աստիճանի հայցման ատենախոսության

ՄԵՂՄԱԳԻՐ

Երևան – 2014

ИНСТИТУТ ПРОБЛЕМ ИНФОРМАТИКИ И АВТОМАТИЗАЦИИ НАН РА

Топчян Тигран Рубенович

ЗАДАЧИ ОПТИМИЗАЦИИ БИЗНЕС ПРОЦЕССОВ НА ПРИМЕРЕ СИСТЕМЫ
ОБРАЗОВАНИЯ

АВТОРЕФЕРАТ

диссертации на соискание ученой степени кандидата технических наук по специальности

05.13.04 – «Математическое и программное обеспечение математических машин,
комплексов, систем и сетей»

Ереван - 2014

Ատենախոսության թեման հաստատվել է Երևանի պետական համալսարանում

Գիտական ղեկավար՝ ֆիզ.մաթ.գիտ. դոկտոր Ս.Կ. Շուքուրյան

Պաշտոնական ընդդիմախոսներ՝ ֆիզ.մաթ.գիտ.դոկտոր Լ.Հ. Ասլանյան
տեխ.գիտ.թեկնածու Գ.Բ. Մարգարով

Առաջատար կազմակերպություն՝ ՌԳԱ Համակարգային ծրագրավորման
ինստիտուտ

Պաշտպանությունը կայանալու է 2014թ. հունիսի 10-ին, ժ. 16:00-ին ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտում գործող 037 «Ինֆորմատիկա և հաշվողական համակարգեր» մասնագիտական խորհրդի նիստում հետևյալ հասցեով՝ Երևան, 0014, Պ. Սևակի 1:

Ատենախոսությանը կարելի է ծանոթանալ ՀՀ ԳԱԱ ԻԱՊԻ գրադարանում:
Սեղմագիրը առաքված է 2014թ. մայիսի 12-ին:

Մասնագիտական խորհրդի
գիտական քարտուղար, ֆ.մ.գ.դ.



Հ. Գ. Սարգսյանյան

Тема диссертации утверждена в Ереванском государственном университете
Научный руководитель: доктор физ.-мат. наук С. К. Шукурян

Официальные оппоненты: доктор физ.-мат. наук Л.А. Асланян
кандидат тех. наук Г.И. Маргаров

Ведущая организация: Институт системного программирования РАН

Защита состоится 10-ого июня 2014г. в 16:00 на заседании специализированного совета 037 «Информатика и вычислительные системы» Института проблем информатики и автоматизации НАН РА по адресу: 0014, г. Ереван, ул. П. Севака 1.

С диссертацией можно ознакомиться в библиотеке ИПИА НАН РА.

Автореферат разослан 12-ого мая 2014г.

Ученый секретарь специализированного
совета, д.ф.м.н.



А. Г. Саруханян

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы. Основной целью системы образования является обеспечение общества специалистами, обладающими набором востребованных компетенций¹. Особого внимания заслуживают программы специальностей высшего образования, ориентированные на использование и разработку технологий, как одни из наиболее динамически изменяющихся в смысле частоты обновления и необходимости кардинального пересмотра учебных программ². Ключевой здесь является задача периодической эффективной настройки процессов подготовки специалистов к требованиям работодателей и обеспечения рынка необходимыми квалифицированными кадрами.

Для указанной оптимизации процессов обучения необходимо модифицировать методы информационного взаимодействия между работодателями и участниками образовательного процесса³. Взаимодействие должно иметь адаптивный характер и обеспечивать достаточную чувствительность к динамике изменения рынка.

Каким бы путем не осуществлялось упомянутое взаимодействие, учебное заведение является конечным исполнителем действий по повышению эффективности процессов обучения и оценки целесообразности изменения учебных программ, вследствие чего в большинстве случаев оно и является инициатором и аккумулятором необходимой информации.

В первую очередь учебные заведения учитывают требования рынка, описанные в различных квалификационных рамках. Рамки квалификаций, создаваемые экспертами области, представляют обобщенные требования, но при этом не отражают динамику области. Примерами рамок являются рамка Задумка-Проект-Реализация-Эксплуатация (CDIO) Массачусетского технологического университета, европейская рамка ИКТ-Компетенций (e-CF) и т.д.

Учебные заведения используют также данные, полученные методами фокус групп и опросов. Однако высокая стоимость организации опросов, низкая мотивация опрашиваемых экспертов и необходимость повторных опросов в случае, если необходим мониторинг динамических изменений во времени затрудняет их применение. Следует также отметить, что задача организации опросов сильно усложняется в условиях глобализации рынка труда.

Другим источником информации для усовершенствования учебных программ являются должностные обязанности (компетенции), описываемые в вакансиях работодателей⁴. Вакансии, в интегральном смысле, обладают свойством адаптивно-

¹DeSeCo. (2005) *Definition and selection of competences: Theoretical and conceptual foundations*. OECD.

²Crawley, E. F., Malmqvist, J., Östlund, S., Brodeur, D. R., *Rethinking Engineering Education: The CDIO Approach*, Springer, New York, 2007.

³O. V. Boev, I. Freeston, G. Heitmann, E. S. Kulyukina, M. S. Tayurskaya, *Engineering Curriculum Design Aligned with Accreditation Standards*, Hochschule Wismar, 2013

⁴Malzahn, N., Ziebarth, S., & Hoppe, H. U. (2013). *Semi-automatic creation and exploitation of competence ontologies for trend aware profiling, matching and planning*. Knowledge Management & E-Learning, 5(1), 84–103

сти к динамике рынка. Работодатель мотивирован в предоставлении актуальной информации.

Перечисленные преимущества, делают данный подход очень привлекательным, однако, большой объем и разнородность данных затрудняет их прямое использование учебными заведениями без предварительной обработки, систематизации и построения информационной модели требований рынка. Построение такой модели позволит компетентными подразделениями учебных заведений анализировать свойства рынка, принять их во внимание и нацелиться на повышение эффективности учебных программ. Модель должна отражать полный спектр компетенций, требуемых рынком, быть чувствительной и адаптивной к динамике изменений. Для облегчения работы экспертов образования необходимо также построить инструментальные средства, позволяющие онлайн аккумуляцию вакансий, их обработку, настройку модели. Они должны предоставлять возможность как количественного, так и качественного анализа для принятия решений по усовершенствованию образовательных программ.

Целью работы является обоснование, проектирование, реализация и экспериментальная апробация программной системы мониторинга данных о компетенциях, требуемых работодателями, их анализа на основе вероятностной модели описания трудовых квалификаций и выдача результатов анализа в виде рекомендаций, служащих исходной информацией для принятия решений по оптимизации образовательного процесса на основе п с работодателями.

Методы исследования, используемые в диссертационной работе, включают методы моделирования, основанные на вероятностных моделях, методы статистической обработки данных и методы разработки программного обеспечения.

Научная новизна. В диссертационной работе были получены следующие результаты:

- Построена и реализована инфраструктура мониторинга данных о компетенциях, их анализа и выдачи результатов анализа в виде рекомендаций. На основе данной инфраструктуры могут быть построены методики усовершенствования образовательных процессов в высших учебных заведениях с учетом обратной связи от работодателей.
- Разработан метод захвата, обработки и хранения больших объёмов вакансий на работу, представленных в интернет среде.
- Предложен новый подход для анализа и классификации компетенций, востребованных работодателями. Подход основан на построении вероятностной модели компетенций по множеству описаний трудовых обязанностей, представленных в онлайн объявлениях на замещение вакантных позиций.
- Предложен метод сравнительного анализа существующих форм описания рамок инженерных компетенций и результатов, полученных с помощью построенной вероятностной модели компетенций.

- Предложен способ определения динамики изменений компетенций, востребованных работодателями, и построения рекомендаций по использованию полученных данных в образовательных процессах. Для упрощения качественного анализа рынка труда и для удобства преподавательского состава учебного заведения разработаны средства визуализации динамики изменений требований рынка к трудовым качествам работников.

Практическое значение. Разработанная программная система использована при реализации проекта ARARAT, выполняемого в рамках программы TEMPUS Европейского Сообщества. Предложенная в работе вероятностная модель используется для сбора информации о текущих тенденциях квалификационных требований в отраслях информационных технологий, финансов и инженерии в целом. Результаты по ее использованию показали, что предлагаемая система протестирована, удобна для использования и может быть применена для оптимизации бизнес процессов планирования разных компонентов учебного процесса как в области инженерного образования в целом, так и, в частности, для образовательных программ в области информационных технологий.

Реализованная программная система позволяет работать с несколькими языками и проверена для англо-армяно-русско язычных онлайн источников вакансий.

Положения, выносимые на защиту:

- Инфраструктура мониторинга данных о компетенциях, их анализа и выдачи результатов в виде рекомендаций.
- Метод построения вероятностной тематической модели описания компетенций по заданному множеству описаний трудовых обязанностей, представленных в свободном формате на сайтах онлайн вакансий.
- Метод сравнительного анализа описаний образовательных стандартов CDIO и e-CF с построенной вероятностной моделью описания трудовых компетенций, который может быть использован для выработки рекомендации по построению методик усовершенствования учебных процессов в инженерных высших учебных заведениях.
- Метод анализа данных, основывающийся на вероятностной тематической модели с временной характеристикой входного потока данных.
- Алгоритм анализа временной динамики описателей компетенций и визуализация результатов, облегчающая их качественную оценку.
- Актуальные характеристики компетенций востребованных на рынке труда специалистов в области разработки и эксплуатации информационных систем.

На основе результатов диссертационной работы создан пакет онлайн сервисов, позволяющий внедрять компоненты в существующие системы автоматизации бизнес процессов образовательных учреждений с целью их оптимизации и повышения эффективности. Пакет реализован в среде Python.

Апробация и публикации работы. Основные результаты и положения диссертационной работы обсуждались и докладывались на семинарах кафедры «Информационных систем» ЕГУ (2012-2014 гг.), общих семинарах Образовательного и исследовательского центра информационных технологий ЕГУ, Института проблем информатики и автоматизации НАН РА, представлялись на международной конференции по вычислительным наукам и информационным технологиям “Computer Sciences and Information Technologies” (CSIT) 2013 в г. Ереване, Армения и на международной научно-практической конференции «Актуальные проблемы подготовки специалистов ИКТ», в г. Хмельницкий, Украина, 15-19 Май 2013.

Результаты диссертации обсуждались также на регулярных встречах и семинарах рабочей группы международного проекта ARARAT, выполняемого в рамках программы TEMPUS Европейского Сообщества.

Результаты работы отражены в пяти публикациях, список которых приведен в конце автореферата.

Структура и объем работы. Диссертация состоит из введения, шести глав, заключения и списка использованной литературы (70 наименований). Основной текст изложен на 115-и страницах.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность работы и практическая значимость темы диссертационной работы, кратко изложено состояние предметной области, сформулированы цели и основные задачи исследования, выделены научные результаты, отличающиеся новизной, положения, выносимые на защиту, и приведена практическая ценность полученных результатов.

В первой главе рассмотрены задачи взаимодействия образовательных учреждений и работодателей с целью обеспечения подготовки кадров, обладающих компетенциями, востребованными рынком труда.

В параграфе 1.1 проанализированы преимущества и недостатки основных трех источников данных, содержащих требования рынка: рамки квалификаций, опросы участников образовательного процесса и вакансии на работу. Подчеркнуты преимущества оптимизации учебного процесса на основе анализа динамического потока онлайн объявлений о вакансиях. Обоснована целесообразность построения вероятностной информационной модели описания корпуса вакансий, обладающей адаптивными свойствами к изменяющимся требованиям рынка. Информационная модель должна отражать тематический состав компетенций, востребованных в корпусе вакансий, и должна выделять часто используемые трафареты совместно используемых компетенций. Эти трафареты могут быть использованы для экс-

пертного анализа и принятие решений по усовершенствованию программ образовательного заведения.

В параграфе 1.2 более детально рассмотрен аппарат рамок квалификаций. В частности, представлены рамка квалификаций инженерного образования CDIO и европейская рамка ИКТ компетенций (e-CF).

В параграфе 1.3 в качестве основной задачи диссертационной работы ставится автоматизация рабочего места эксперта образовательных систем. Система автоматизации должна обеспечивать сбор и обработку вакансий из различных онлайн источников размещения вакансий, оперативное построение модели, измерение различных характеристик, отражающих изменение требований рынка труда и визуализацию динамики изменений. Система должна обеспечивать анализ корреляции данных модели с известными моделями требований рынка, таких как рамки квалификаций и примерные образовательные программы. Предполагается, что преимуществом системы должен быть принцип ее построения, основанный на методах машинного обучения.

Во второй главе диссертации проводится анализ различных методов, применяемых для выделения семантики из большого корпуса немаркированных документов, и обосновывается применение вероятностной тематической модели для решения поставленной задачи методами машинного обучения⁵

В параграфе 2.1 проведен анализ характеристик информационной модели, необходимых для решения поставленной задачи.

В параграфе 2.2 рассмотрен важный подкласс ориентированных вероятностных тематических моделей⁶ (directed probabilistic topic models, DPTM), которые осуществляют мягкую кластеризацию и применяются для выделения тематики текстов в больших коллекциях документов.

В третьем параграфе обосновано применение вероятностных тематических моделей к решению задачи анализа вакансий и построения тематической модели компетенций. Вероятностное тематическое моделирование⁷ — это набор алгоритмов, которые применяются для выявления неявных/скрытых тематических структур из большого количества документов. Алгоритмы тематического моделирования основаны на статистических методах. Они анализируют слова из исходных документов для выявления скрытых тем присутствующих в общем наборе документов, связей между этими темами и изменения тем во времени. Данные алгоритмы не требуют какой-либо предварительной аннотации или маркировки документов: темы выявляются в результате анализа исходных документов. Результаты алгоритмов тематического моделирования могут быть использованы для исследования,

⁵I. Witten, E. Frank, and M. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Mateo, CA, 3rd edition, 2011.

⁶Ali Daud, Juanzi Li, Lizhu Zhou, Faqir Muhammad, *Knowledge discovery through directed probabilistic topic models: a survey*, Frontiers of Computer Science in China, June 2010, Volume 4, Issue 2, pp 280-301

⁷D. Blei. *Probabilistic topic models*. Communications of the ACM, 55(4):77–84, 2012.

linux	software	experience	business	windows
servers	based	grow	requirements	systems
install	experience	company	functional	server
plus	development	work	analyst	exchange
architect	paid	salary	technical	experience
infrastructure	leading	please	create	administrator
monitoring	small	full	systems	microsoft
tools	expertise	benefits	design	active
various	daily	start	user	directory
services	developers	time	support	administration

Рис. 1: Пример пяти тем полученных из корпуса вакансий для ИТ сектора. Показана десятка самых популярных слов.

обобщения, визуализации и построения теоритических моделей в рамках определенного набора данных.

В третьей главе диссертации представлено применение вероятностного тематического моделирования к задаче анализа компетенций, характеризующих рынок труда.

В параграфе 3.1 приведено описание тематической модели латентного размещения Дирихле⁸ (LDA). Данная модель является порождающей вероятностной моделью дискретных данных, таких как коллекции документов о вакансиях. LDA моделирует каждый элемент коллекции как распределение по множеству «тем» (topics). В свою очередь каждая тема моделируется как распределение по фиксированному словарю. Особенностью LDA является то, что этот метод не учитывает порядок слов в документе и порядок документов в коллекции, а работает с частотой слов и пытается вывести ее путем помещения контекстуально взаимосвязанных слов в одну тему. В свою очередь, данная особенность является также и сильной стороной, так как модель может быть применена к любым текстовым данным почти на любом языке без особых изменений. Еще одним предположением в модели LDA является то, что число тем известно и зафиксировано заранее. На Рис. 1 изображены пять тем, выделенных в результате выполненного нами анализа корпуса вакансий для ИКТ сектора на основе онлайн источника www.indeed.com.

В параграфе 3.2 выполнена адаптация метода LDA для получения тематической модели компетенций. Основным здесь является моделирование коллекции данных с использованием K тем компетенций. Каждая из K тем будет содержать

⁸D. Blei, A. Ng, and M. Jordan. *Latent Dirichlet allocation*. Journal of Machine Learning Research, 3:993–1022, January 2003

множество слов с различными частотами и каждый документ будет иметь различные частоты для каждой темы. Это будет означать, что каждое описание вакансий может быть декомпозировано в конечный набор тем, которые в свою очередь будут описывать суть, представленную в документе. Например, описание вакансии может быть смоделировано как описание требований относительно компетенций «опыт работы» и «навыки самостоятельной работы», так и описание требований относительно компетенций «опыт» и «коммуникационные способности». Обобщая рассмотренный пример можно сказать, что с каждой темой связывается фиксированный словарь терминов свойственных области и каждая тема позволяет описывать область как в широком, так и в узком смысле. Таким образом LDA выделяет гетерогенность описаний вакансий, используя модель смешанной принадлежности, что очень важно для адекватного представления корпуса данных.

Заметим, что заранее известно только количество тем (K), а сами темы выделяются в результате порождающего процесса LDA. Тем самым LDA выявляет скрытые закономерности, наиболее подходящие исходным данным. В нашем случае выделяются компетенции из описания вакансий.

Для описания порождающего процесса введем следующие обозначения:

Обозначение	Описание
K	данное количество тем.
V	упорядоченное множество всех уникальных слов из корпуса данных. Называется словарем. Мощность множества V обозначим \bar{V}
D	множество документов. С каждым документом связывается \bar{V} -мерный вектор, компонент i которого показывает сколько раз слово i из V встречается в данном документе.
d	элемент множества документов D
θ_d	K -мерный вектор пропорций тем для документа d .
β_k	\bar{V} -мерный вектор пропорций слов для темы с индексом k
$z_{d,n}$	назначение тем для слова с индексом n из документа d .
z_d	назначение тем для слов из документа d .

Порождающий процесс описывается следующим образом:

1. Для каждой темы k из исходных K тем:
 - (a) Получить распределение β_k над словарем V .
2. Для каждого документа d :
 - (a) Получить распределение θ_d .
 - (b) Для каждого слова w с индексом n из документа d :
 - i. Выбрать распределение $z_{d,n}$ из θ_d , где каждое $z_{d,n} \in [1, K]$

ii. Выбрать слово $w_{d,n}$ из $\beta_{z_d,n}$

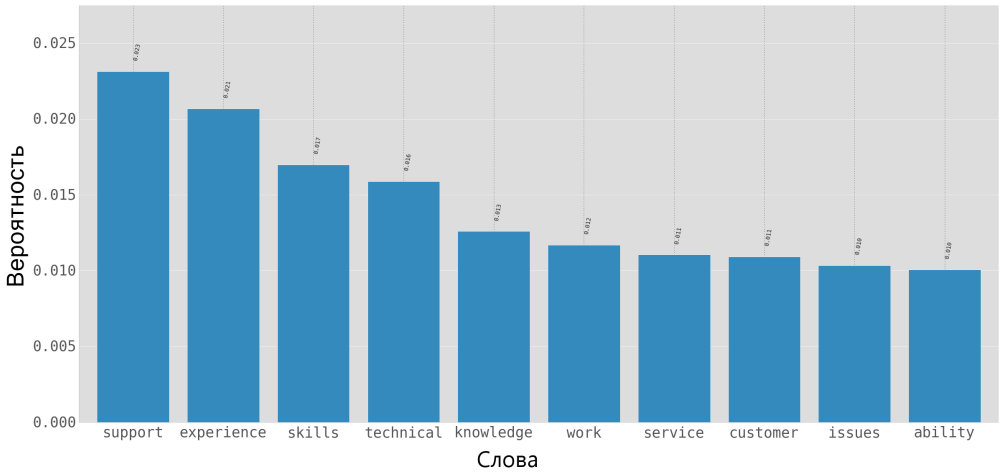


Рис. 2: Десятка самых вероятных слов для наиболее часто встречающейся темы из корпуса данных ИКТ.

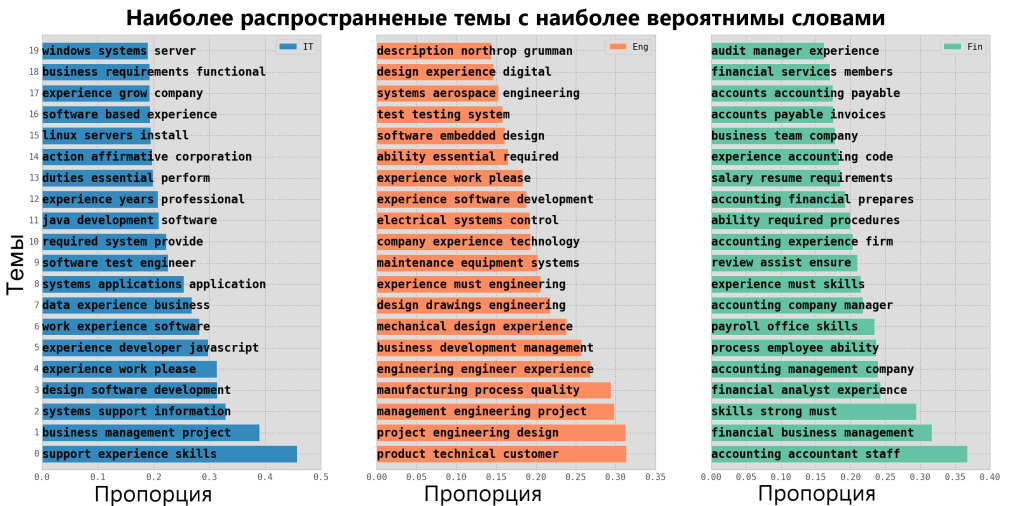


Рис. 3: Двадцатка наиболее часто встречающихся тем. Представлены, также, три наиболее часто встречающихся слова для каждой темы.

где β_k и θ_d моделируются, используя распределение Дирихле. Распределение тем по документам и распределение слов по темам задаются полиномиальным распределением, определённым над θ_d и β_{z_d} , соответственно.

Далее будут представлены результаты обучения модели LDA для различного числа тем на корпусе описаний вакансий, состоящем из 10,000 документов, и будут предложены методы для анализа и вывода результатов из данных и модели.

В результате обучения модели получается объединённое распределение над наблюдаемыми и скрытыми случайными переменными. Затем структура тем выводится из апостериорного распределения скрытых переменных, в рассматриваемом корпусе документов. Прямое вычисление апостериорного распределения не целесообразно ввиду большой сложности расчётов. С другой стороны, необходимое апостериорное ожидание скрытых переменных можно получить методом вариационного исчисления.

Пропорции тем β_k , пропорции тем для документов θ_d и принадлежность слов к темам z_d позволяют в полной мере исследовать корпус данных. Например, мы можем использовать их для анализа отдельных тем, извлечения пропорционального распределения в рамках корпуса, нахождения наиболее часто повторяющихся тем, проецирования новых документов в пространство тем и кластеризации документов, основываясь на их тематическом распределении.

Для более детального исследования тематической структуры корпуса описаний вакансий предлагаются три подхода.

Анализ отдельных тем. Мы можем использовать β_k и $z_{d,n}$ для анализа отдельных тем (Рис. 2). Каждое слово из z_d будет встречаться в теме с определенной



Рис. 4: Тематическая структура документа. Показаны пять самых вероятных слов для данной темы.

частотой из β_k . Некоторые слова будут вероятнее других и рассматривая некоторые из самых вероятных слов мы можем понять суть темы. По существу, декомпозиция тем подчеркивает насколько часто данные группы слов встречаются вместе, что предоставляет ценную информацию об анализируемых описания вакансий.

Нахождение наилучших тем методом упорядочивания. Одна из главных задач – это нахождение тем, которые предоставляют ценную информацию об анализируемых данные. В случае большого числа тем, ручное просеивание не желательно и естественное упорядочение тем – это упорядочение по популярности. Для нахождения тем, которые встречаются в большинстве документов необходимо использовать частоты из θ_d (Рис. 3).

Анализ по отдельным описаниям вакансий. Мы также можем использовать β_k , θ_d и z_d для проецирования, любого существующего или нового документа в пространство тем обученной модели. То есть, мы можем анализировать описание одной вакансии и получить соответствующее распределение тем. Например, имея анализ описания вакансии мы может получить соответствующую тематическую структуру (Рис. 4).

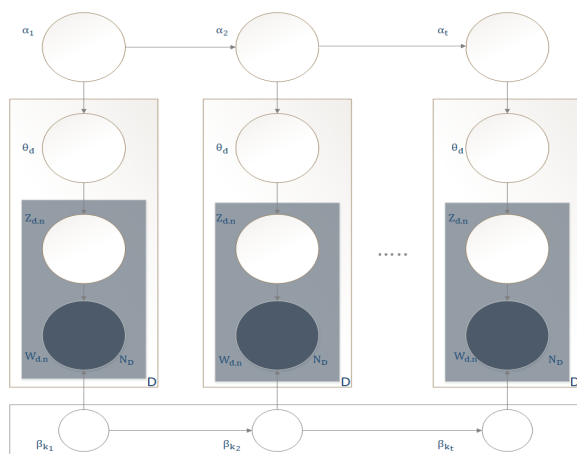


Рис. 5: Графическая модель для DTM.

В четвертой главе диссертации представлен метод анализа динамики компетенций на основе динамических тематических моделей. Рынок труда имеет динамику развития. Возникают новые компетенции и квалификации, изменяются приоритеты существующих. Анализ этой динамики очень важен для всестороннего понимания требований индустрии и совершенствования образовательных программ⁹.

⁹Crawley, E. F., Malmqvist, J., Östlund, S., Brodeur, D. R., *Rethinking Engineering Education: The CDIO Approach*, Springer, New York, 2007.

В параграфе 4.1 описаны динамические тематические модели, которые учитывают дополнительную информацию о времени создания или публикации вакансий. Эти модели рассматривают коллекцию документов отдельными срезами во времени. Определенные временные метки могут быть представлены любыми, равномерно распределёнными, деноминациями (часы, дни, месяцы, годы). В отличие от статической модели LDA, где K тем были описаны как распределение Дирихле над фиксированным словарем слов V , динамические тематические модели взаимосвязывают параметры тем в модель пространства состояний.

Порождающий процесс подобен порождающему процессу LDA, но параметры тем и документов взаимосвязаны. Этот процесс представлен в виде графической модели на Рис.5.

В параграфе 4.2 описан анализ динамики компетенции. Динамическая тематическая модель применена для анализа корпуса вакансий в ИКТ индустрии.

В контексте динамического анализа наиболее интересны изменения выделенных тем во времени. Рассматривается изменение вероятности каждого слова темы во времени. На Рис.6 динамика представлена графиком изменений вероятностей во времени для четырех самых популярных слов. Изменение темы представлено в виде облака слов, где размер слова в облаке прямо пропорционален вероятностной составляющей в теме.

Заметим, что на Рис. 6 представлен результат для темы, прямо коррелирующей с группой компетенций "Test, Verification, Validation and Certification" из рамки CDIO. Можно заметить, что вероятности для приведенных компетенций достаточно стабильны во времени. Это можно объяснить тем, что данные компетенции встречаются в большом числе должностных описаний. Это можно считать воодушевляющим примером который, в свою очередь, доказывает полезность модели для выявления ценной информации из набора данных.

В пятой главе диссертации предлагается модель, для сравнительного анализа существующих форм описания рамок компетенций (например, рамок квалификаций CDIO и e-CF) и результатов, полученных с помощью вероятностного тематического моделирования.

В параграфе 5.1 дается подробное описание модели. В данной модели рамка компетенций задается разбиением на области компетенций. В свою очередь, каждая область рамки уникальным образом разделена на группы компетенций, которые представлены ключевыми словами, выделенными из исходного текстового описания рамки. Предлагаемая модель основана на модели LDA и определяет отображение каждой темы на области компетенций рамки. Отображение строится на основе соответствия компетенции темы и ключевых слов области рамки.

В результате построения отображения каждая компетенция темы получает маркировку приоритетной принадлежности к одной из областей рамки. Маркируя каждую компетенцию в контексте рамки компетенций, мы не только можем найдем маркировку тем в целом, а также найдем псевдо-вероятность принадлежности темы к данной компетенции. Следовательно, для темы β_k :

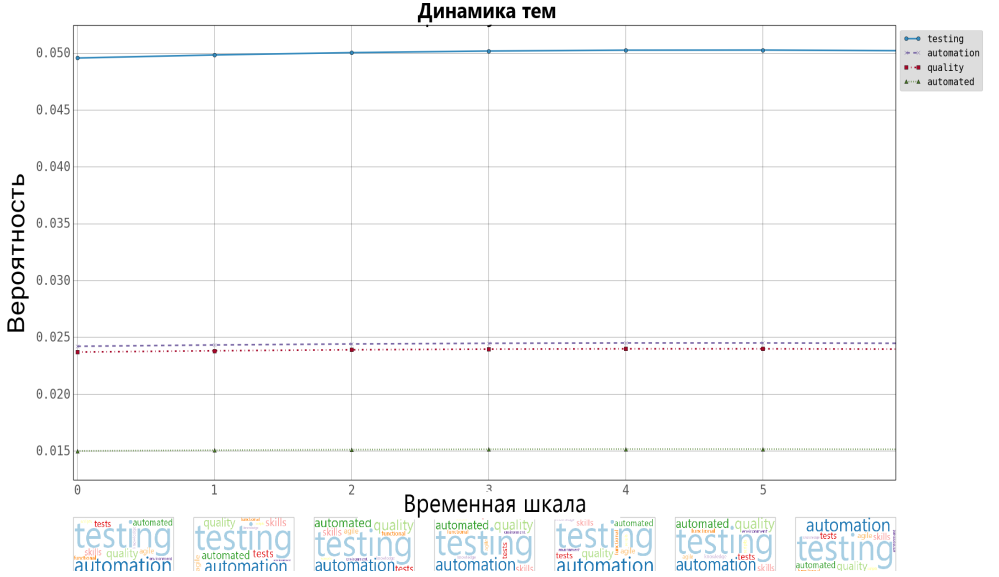


Рис. 6: Динамика тем соответствующих компетенции 'IMPLEMENTING' из CDIO.

$$\beta_k = [(w_1, p_1), \dots, (w_l, p_l)] \longleftrightarrow C_k = [(c_{k_1}, p_1^*), \dots, (c_{k_{areas}}, p_{areas}^*)] \text{ for } l \in [0, n] \quad (1)$$

Где C_k - это вектор пар области компетенций (c) и их интегральной пропорции (p^*).

Для нахождения вектора C_k нам необходима метрика для смыслового сравнения слов. В данной работе рассмотрены шесть различных метрик, основанных на графе WordNet.

Для каждого слова w нахождение его области компетенций c_w возможно следующим образом:

$$c_w = \max_{\substack{i \in \\ (0, areas)}} \left(\frac{\sum_{j=1}^p wnsim(w, c_{ij})}{p} \right) \quad (2)$$

В уравнении 2 $wnsim$ - это функция подобности из числа рассмотренных шести метрик. Используя уравнение 2 возможно найти область компетенций каждого слова ($c_{kw_1}, \dots, c_{kw_l}$) для темы k и используя эту вероятность области компетенций c_i выводится следующей формулой:

$$p_{c_i}^* = \left(\text{norm} \left(\sum p_j \right) \mid \forall j : c_{k_{w_j}} = c_i \right) \quad (3)$$

Таким образом можно найти $(p_1^*, \dots, p_{areas}^*)$.

Таким образом выводится векторное представление C_k из уравнения 1. Аналогичный подход можно применить к другим рамкам компетенций, например, e-CF.

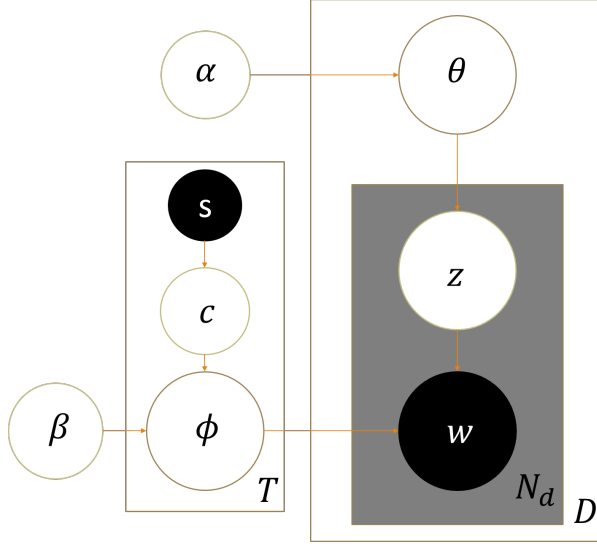


Рис. 7: Графическое представление предлагаемой модели.

На основе определённого отображения предлагаемая модель может быть представлена как графическая модель на Рис.7. Порождающий процесс данной модели можно описать следующим методом:

1. Для каждой темы k из исходных K тем:
 - (a) Получить распределение β_k над словарем V .
2. Для каждого документа d :
 - (a) Получить распределение θ_d .
 - (b) Для каждого слова w с индексом n из документа d :
 - i. Выбрать распределение $z_{d,n}$ из θ_d , где каждое $z_{d,n} \in [1, K]$
 - ii. Выбрать слово $w_{d,n}$ из $\beta_{z_{d,n}}$
3. Для каждой темы k из K



Рис. 8: Пропорций областей компетенций для 120 тем

- Для каждого слова w создать назначение c_w из s
- Для каждой области a $p_{k,a}$ для β_k и $c_{k,w}$

В параграфе 5.2 описаны результаты применения предлагаемой модели на корпусе вакансий и рамок компетенций CDIO и e-CF. Представлены результаты для всех шести рассмотренных метрик. В качестве примера, на Рис.8 представлены пропорции областей компетенций для 120 тем, выделенных из корпуса вакансий, а на Рис.9 изображены ближайшие темы для отдельных областей.

В шестой главе диссертации описывается метод сбора и обработки описаний вакансий из множества источников. Для построения полноценного представления о рынке труда и компетенциях необходима соответствующая инфраструктура захвата и обработки данных. Эта инфраструктура должна поддерживать следующие возможности:

- сбор данных из множества источников;
- непрерывность сбора и онлайн пополнение данных;
- масштабируемость захвата и обработки данных;
- модульная поддержка разнообразных методов анализа данных;
- наличие протоколов для быстрого доступа, обработки, визуализации и анализа данных

Ближайшая тема для данной области

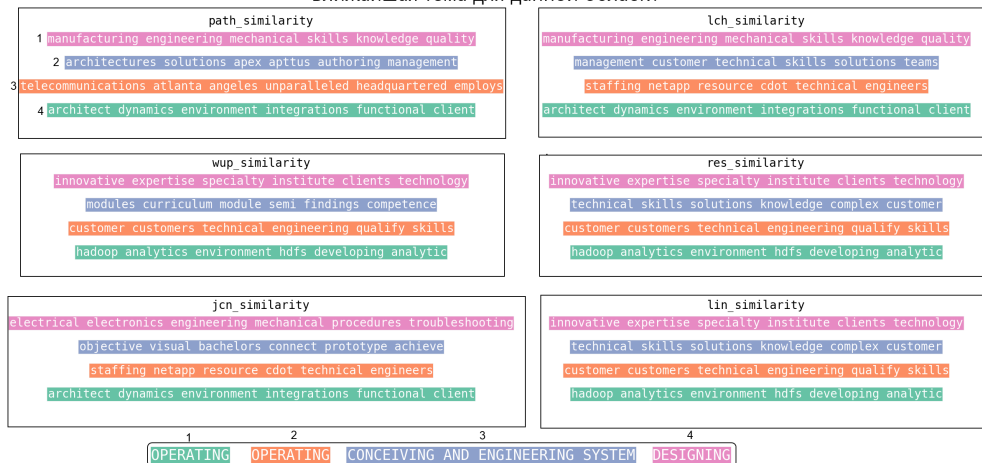


Рис. 9: Ближайшие темы для определенной области

Для достижения выше описанных возможностей была создана программная система, выполненная в классической трех уровневой архитектуре: уровень данных, сервисный уровень и презентационный уровень. На Рис.10 представлена принципиальная схема системы.

В параграфе 6.1 описывается уровень данных системы. Детально описывается модуль аккумуляции вакансий из онлайн источников, модуль обработки данных и модуль хранения информации.

В параграфе 6.2 описывается сервисный уровень системы. В частности, описываются модули реализующие модели, описанные в предыдущих главах. Также представлена реализация веб сервисов, реализующих интерфейс с другими уровнями архитектуры.

В параграфе 6.3 описана реализация пользовательского интерфейса для инструментария анализа и мониторинга данных о компетенциях и выдачи результатов анализа в виде рекомендаций.

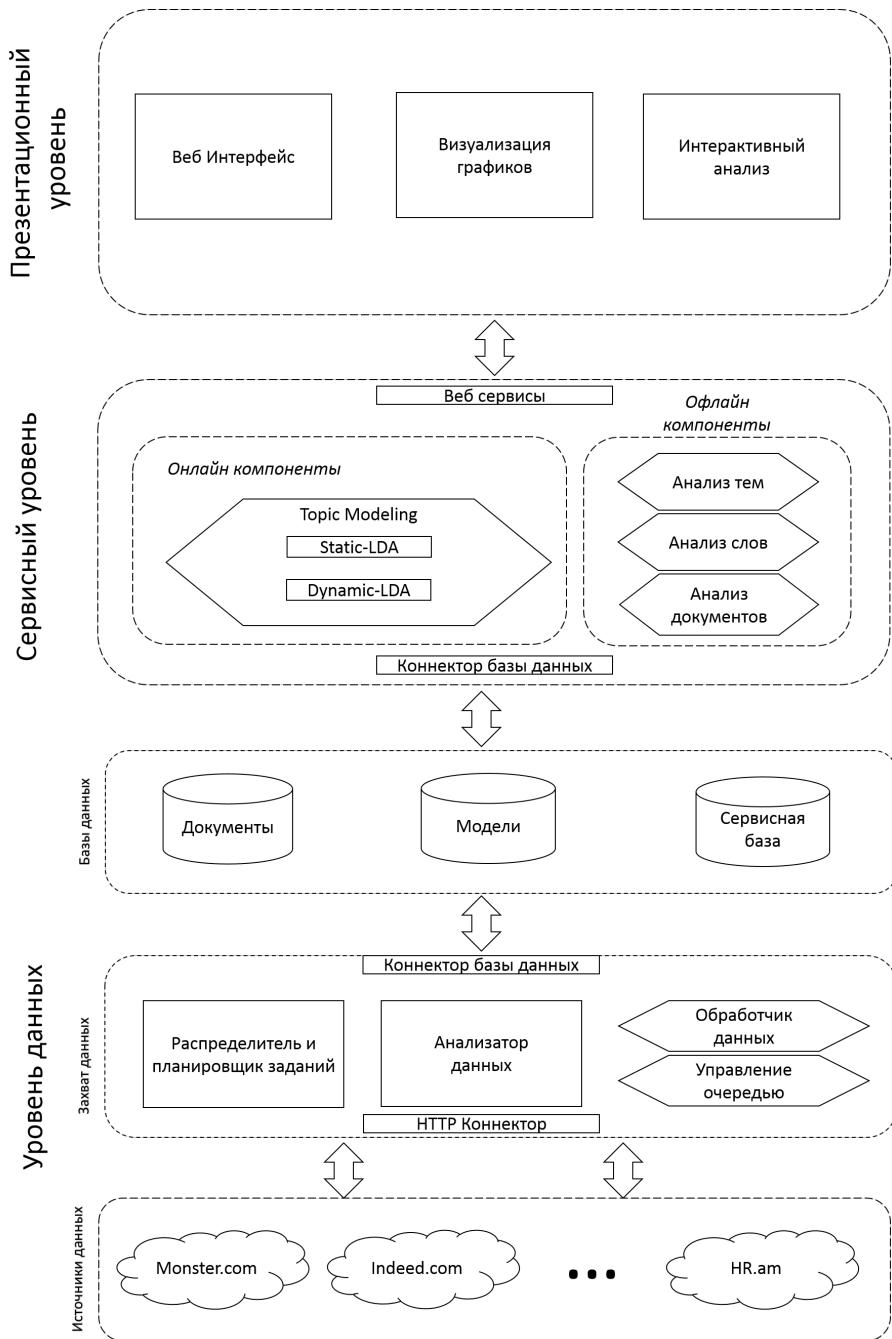


Рис. 10: Принципиальная схема системы.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ДИЦЕРТАЦИОННОЙ РАБОТЫ

- Инфраструктура мониторинга данных о компетенциях, их анализа и выдачи результатов в виде рекомендаций.
- Метод построения вероятностной тематической модели описания компетенций по заданному множеству описаний трудовых обязанностей, представленных в свободном формате на сайтах онлайн вакансий.
- Метод сравнительного анализа описаний образовательных стандартов CDIO и e-CF с построенной вероятностной моделью описания трудовых компетенций, который может быть использован для выработки рекомендации по построению методик усовершенствования учебных процессов в инженерных высших учебных заведениях.
- Метод анализа данных, основывающийся на вероятностной тематической модели с временной характеристикой входного потока данных.
- Алгоритм анализа временной динамики описателей компетенций и визуализация результатов, облегчающая их качественную оценку.
- Актуальные характеристики компетенций, востребованных на рынке труда, специалистов в области разработки и эксплуатации информационных систем.

СПИСОК РАБОТ ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

- [1] A. Topchyan and T. Topchyan. Muscle-based skeletal bipedal locomotion using neural evolution. In *Proceedings of IEEE, Computer Science and Information Technologies (CSIT), 2013*, pages 1–6, Sept 2013.
- [2] A. Topchyan and T. Topchyan. Muscle-based skeletal bipedal locomotion using neural evolution. *Computer Science and Information Technologies (CSIT)*, 2013.
- [3] Tigran Topchyan. Engineering competence frameworks and topic modelling. *Mathematical Problems of Computer Science*, 2014.
- [4] Tigran Topchyan. Job market requirements and qualities extraction for qualification. *Mathematical Problems of Computer Science*, 2014.
- [5] Tigran Topchyan. Market-required competence topic dynamics. *Classification, Forecasting, Data Mining (CFDM)*, 2014.

Resume

Tigran Topchyan

BUSINESS PROCESS OPTIMIZATION IN THE CONTEXT OF EDUCATION SYSTEMS

The main goal of an education system is creation of specialists possessing a set of sought after competences and qualification. Particularly noteworthy are fields focused on technology operation and development, as one of the most dynamic in the sense of changes in the field and the need for a radical revision of the curricula. The key problem is the effective periodic customization of the processes for educating specialists and the provision of required skills to the market. For this kind of optimization it is necessary to modify of the methods of informational interaction between employers and participants in the educational process. The interactions must be adaptive in nature and be sensitive to the dynamics of changes in the market.

Whichever way the mentioned interactions takes place, the educational institution is ultimately responsible for improving the effectiveness of education processes and the assessment of the curriculum change feasibility, so in most cases it is the initiator and recipient of the information.

First of all educational institutions take into account, the requirements of the market, which are described in various qualifications frameworks. Qualifications Frameworks are created by industry experts, but nevertheless do not reflect the dynamics of the area. Examples of these frameworks are Conceive — Design — Implement — Operate (CDIO) and the European e-Competency Frameworks (e-CF).

Educational institutions also use data obtained by means of focus groups and surveys. However, the high organizational costs, the lack of motivation from the experts being surveyed and the need for repeated surveys to monitor the dynamics and trends of the time, creates problem to carrying them out. It is also worth noting that the task of organizing the surveys is much more complicated in the context of the globalization of the labour market. Another source for the improvement of educational programs are the job requirements (competencies), described in job listings. The descriptions are built based on the competencies required by employers.

The job descriptions, in the integral sense, are adaptive to the dynamics of the labor market. The employers are motivated to provide the relevant information.

The supplied advantages make this an attractive approach, but the large volume and the diversity of the data, makes it difficult for educational institutions to use them without preprocessing, systematization and the creation of an informational model of the market. The creation of such a model would allow the competent departments of these institutions to analyse the properties of the market, take them into account and aim at improving the effectiveness of educational programs. The model must reflect the full spectrum of competencies, required by the market, be sensitive and adaptive to the dynamics of change. To facilitate the work of education experts, the creation of a toolset

for online acquisition of job descriptions, their processing and the creation of a model, is necessary. They should conduct periodic analysis of data to identify the dynamics of the labor market requirements and provide the ability to provide qualitative conclusions for decision-making and the improvement of educational programs.

Main results of the work

- Infrastructure for data monitoring of competencies, their analysis and the delivery of results of analysis in the form of recommendations;
- A method for comparative analysis of educational standard descriptions (Conceive — Design — Implement — Operate (CDIO) and the European e-Competency Frameworks (e-CF)) based on the created probabilistic model for competencies, which can be used to generate recommendations for the construction of methodologies for the modification and improvement of the educational programs in engineering universities.
- A method for data analysis, based on probabilistic topic models with a temporal characteristic of the input data stream.
- An algorithm for the analysis of the temporal dynamics of competencies required by the market presented on the example of technological specialities, with the possibility of robust visualization to simplify the qualitative analysis.
- Topical characteristics of competences, required by the labor market in the context of the IT industry. Based on the results of the thesis a package of on-line service was created, which allows for effective integration of components into existing system of business process automation systems, with the goal of optimising and improving their efficiency. The software package was created in the Python environment.
- A method for the creation of probabilistic topic models of competencies from a large corpus of job-description data, freely available from numerous on-line job listing directories.

Ամփոփում

Թուփչյան Տիգրան Ռուբենի

ԲԻԶՆԵՍ ՊՐՈՑԵՍՆԵՐԻ ՕՊՏԻՄԻԶՄԱՅԻ ԽՆԴԻՐՆԵՐԸ ԿՐԹԱԿԱՆ ՀԱՄԱԿԱՐԳԻ ՕՐԻՆԱԿԻ ՎՐԱ

Կրթական համակարգի հիմնական նպատակն է հասարակությունը ապահովել այնպիսի աշխատուժով, որը տիրապետում է պահանջված որակավորումների: Հատուկ ուշադրության են արժանի տեխնոլոգիաների մշակմանը և օգտագործմանը միտված մասնագիտությունների կրթական ծրագրերը, որպես առավել հաճախ և արմատապես փոփոխվող կրթական ծրագրեր: Այստեղ արմատական խնդիր է հանդիսանում մասնագետների պատրաստման գործընթացների պարբերական արդյունավետ հարմարեցում գործատուների պահանջներին և շուկայի ապահովումը անհրաժեշտ որակի կադրերով: Ուսուցման գործընթացների նշված օպտիմալացման համար անհրաժեշտ է բարելավել գործատուների և ուսումնական գործընթացների մասնակիցների միջև տեղեկատվական համագործակցությունը: համագործակցությունը պետք է ունենա հարմարվող բնույթ և ապահովի շուկայի փոփոխությունների նկատմամբ բավարար զգայունություն: Ինչպես էլ իրականացվի նշված համագործակցությունը, ուսումնական հաստատությունն է հանդիսանում արդյունավետության բարձրացման գործողությունների վերջնական կատարողը և փոփոխության մասին որոշումների վերջնական կայացնողը, ուստի և ինքն է հանդիսանում անհրաժեշտ տեղեկատվության հավաքագրման նախաձեռնողը և տեղեկատվության կուտակողը: Առաջին հերթին ուսումնական հաստատությունները հաշվի են առնում աշխատաշուկայի պահանջները, որոնք ներկայացված են որակավորումների տարբեր շրջանակներում: Որակավորումների շրջանակները մշակվում են տիրույթի փորձագետների կողմից և ներկայացնում են ընդհանրական պահանջներ, բայց չեն արտացոլում մասնագիտական տիրույթի դինամիկան: Շրջանակի օրինակ են հանդիսանում Մասաչուսեթսի տեխնոլոգիական համալսարանի Մտահղացում-Պրեյեկտում-Իրականացում-Սպասարկում (CDIO) շրջանակը, S2S-կոմպետենցիաների (e-CF) եվրոպական շրջանակը և այլն:

Կրթական հաստատությունները օգտագործում են նաև ֆոկլուս խմբերի և հարցումների մեթոդով ստացված տվյալներ: Հարցումների կազմակերպման բարձր արժեքը, հարցվող փորձագետների ցածր մոտիվացիան և կրկնակի հարցումների անհրաժեշտությունը, երբ ցանկանում ենք մոնիթորինգի ենթարկել փոփոխությունների դինամիկան բարդեցնում են հարցումների կիրառումը: Հարկ է նշել, որ հարցումների կազմակերպումը բարդանում է աշխատաշուկայի գլոբալացման պայմաններում:

Կրթական ծրագրերի բարելավման համար տեղեկատվության մեկ այլ աղբյուր են հանդիսանում աշխատանքային պարտականությունների նկարագրերը, որոնք ներկայացված են գործատուների թափուր պաշտոնների նկարագրություններում: Պաշտոնի

նկարագրերը ինտեգրալ առումով օժտված են աշխատաշուկայի պահանջներին հարմարվելու հատկությամբ: Գործատուն շահագրգիռ է ներկայացնել ակտուալ տեղեկատվություն: Նշված առավելությունները դրձնում են այս մոտեցումը հետաքրքրական, բայց տեղեկատվության մեծ ծավալը և բազմազանությունը դժվարեցնում են ուղղակի օգտագործումը կրթական ծրագրերի բարելավման համար առանց նախնական մշակման, համակարգման և շուկայի պահանջների տեղեկատվական համակարգի կառուցման: Նման համակարգի կառուցումը թույլ կտար ուսումնական հաստատությունների պատասխանատուներին վերլուծել շուկայի հատկությունները և նպատակաուղղվել դեպի ծրագրերի արդյունավետությունը: Մոդելը պետք է արտացոլի շուկայում պահանջվող կոմպետենցիաների ամբողջ սպեկտրը, լինի զգայուն և ադապտիվ փոփոխությունների դինամիկայի նկատմամբ: Կրթության փորձագետների աշխատանքի հեշտացման համար պետք է կառուցվեն գործիքային միջոցներ, որոնք կապահովեն տվյալների առցանց հավաքագրումը, մշակումը և տվյալների մոդելի կառուցումը: Պետք է կառուցվեն գործիքներ շուկայի պահանջների դինամիկայի բացահայտման նպատակով տվյալների պարբերական վերլուծության համար: Անհրաժեշտ է կարողանալ իրականացնել որակական հետևություններ ծրագրերի բարելավման համար որոշումներ կայացնելու նպատակով:

Ատենախոսության հիմնական արդյունքներն են՝

- Կոմպետենցիաների մասին տվյալների վերահսկման, վերլուծության և երաշխավորությունների տեսքով արդյունքների տրամադրման ինֆրակառուցածքը:
- Առցանց կայքերում ազատ ձևաչափով ներկայացված թափուր պաշտոնների նկարագրերի հիման վրա կոմպետենցիաների նկարագրման թեմաթիկ հավանականային մոդելի կառուցման մեթոդը
- CDIO և e-CF կրթական չափորոշիչների և կոմպետենցիաների նկարագրման համար կառուցված թեմաթիկ հավանականային մոդելի համեմատության մեթոդը, որը կարող է օգտագործվել ինժեներական բարձրագույն ուսումնական հաստատություններում կրթական գործընթացների բարելավման նպատակով երաշխավորություններ առաջարկելու մեթոդաբանության կառուցման համար
- Տվյալների մուտքային հոսքի ժամանակային բնութագրիչի առկայության պայմաններում տվյալների վերլուծության մեթոդը, որը հիմնված է հավանականային թեմաթիկ մոդելի գաղափարի վրա
- Կոմպետենցիաների նկարագրիչների ժամանակային դինամիկայի վերլուծության ալգորիթմը, որը կհեշտացնի նկարագրիչների որակական գնահատումը
- Տեղեկատվական համակարգերի մշակման և սպասարկման բնագավառի մասնագետների կոմպետենցիաների արդի բնութագրիչներ, որոնք պահանջված են աշխատաշուկայում



Ծավալը՝ 24 էջ: Տպաքանակը՝ 100 օրինակ:
ՀՀ ԳԱԱ ԻԱՊԻ համակարգչային պոլիգրաֆիայի լաբորատորիա
Երևան, Պ. Սևակի 1