**Федеральное государственное автономное образовательное учреждение высшего образования   
"Национальный исследовательский университет   
"Высшая школа экономики"**

Факультет Компьютерных наук

Департамент анализа данных и искусственного интеллекта

## Рабочая программа дисциплины Научно-исследовательский семинар "Анализ и майнинг данных"

для образовательной программы «Прикладная математика и информатика» направления подготовки 01.03.02. Прикладная математика и информатика

уровень бакалавриат

Разработчик программы

Осадчий А.Е., кандидат тех. наук, доцент, aossadtchi@hse.ru

Одобрена на заседании департамента анализа данных и искусственного интеллекта

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_ 2015 г.

Руководитель департамента С.О. Кузнецов \_\_\_\_\_\_\_\_

Рекомендована Академическим советом образовательной программы

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_ 2015 г., № протокола\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Утверждена «\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_ 2015 г.

Академический руководитель образовательной программы

А.С. Конушин \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Москва, 2015

*Настоящая программа не может быть использована другими подразделениями университета и другими вузами без разрешения подразделения-разработчика программы.*

# Пояснительная записка

## Авторы программы

кандидат технических наук, доцент А.А. Незнанов;

Ph.d., профессор О.С. Пьяных;

Ph.d., профессор А.Е. Осадчий.

## Требования к студентам

Изучение дисциплины «Научно-исследовательский семинар «Анализ и майнинг данных» на 3-4 курсе требует изучения дисциплины «Научный семинар» на 2 курсе ОП «Прикладная математика и информатика». Знание английского языка на уровне свободного чтения англоязычных текстов желательно.

## Аннотация

Дисциплина «Научный семинар» предназначена для подготовки бакалавров по направлению 010302 – Прикладная математика и информатика.

Основной целью научно-исследовательского семинара «Анализ и майнинг данных» является развитие у бакалавров профессиональной компетенции, необходимой для самостоятельной аналитической работы в прикладных областях, выработка навыков обработки и стандартизации хранения данных, построения модели и эксперимента, а также оформления результатов исследований в виде научных текстов.

Основными навыками и областями компетенции студентов по результатам освоения НИС должны стать:

* доступ к источникам данных (открытым и связанным данным, сетевым хранилищам и специализированным сервисам) в соответствии с международными стандартами на форматы и протоколы передачи данных;
* анализ источников информации – научных публикаций, статистических данных, мультимедийных данных;
* выполнение научных исследований в перспективных прикладных направлениях медицинской информатики при помощи современных методов анализа данных;
* Выбор адекватной модели, идентификация параметров моделей, оценка качества подгонки, выработка рекомендаций по параметрам экспериментального дизайна нейрокогнитивных экспериментов;
* представление результатов исследований (включая научную дискуссию, подготовку научных сообщений, докладов и публикаций в форме коротких статей на студенческую конференцию, публичная презентация проектов).

Вначале, студенты знакомятся с научным подходом и методологией науки в целом и компьютерных наук в частности, благодаря участию всех преподавателей департамента подготавливаются к осознанному выбору тем будущих междисциплинарных и выпускных квалификационных работ, происходит знакомство с основными технологиями и инструментами организации индивидуальной и коллективной научно-исследовательской работы. Получение опыта публичного выступления в рамках описания форматов открытых данных становится важной составляющей вводного блока. Изучаются:

* Методология науки в целом и компьютерных наук в частности.
* Научные направления научных исследований преподавателей факультета компьютерных наук и аффилированных с ним ОП.
* Основы ведения научной дискуссии с приглашением ведущих руководителей секций научных семинаров. Формализация, краткое изложение и визуализация результатов исследований в письменной и устной формах.
* Основы работы в системе компьютерной верстки естественно-научных текстов *TeX* (*LaTeX*).
* Составление списков использованных источников, цитирование и работа с библиографическими базами (*BibTeX*, *BibLaTeX*). Ведение собственных тематических и авторских библиографий с использованием *JabRef*, *Mendeley*, *Zotero*.
* Коллективная работа над исправлением замечаний рецензента в онлайн-сервисе [www.sharelatex.com](http://www.sharelatex.com).
* Подготовка мультимедийной презентации в *Microsoft PowerPoint* или с использованием пакета *Biber* для *LaTeX* документов.
* Дистанционная защита курсовой работы с использованием сервисов [www.gotomeeting.com](http://www.gotomeeting.com), [www.zoom.us](http://www.zoom.us).

В продолжение курса студенты знакомятся с новым для них междисциплинарном направлением исследований – *медицинской информатикой* – одной из наиболее динамично развивающихся дисциплин, возникшей на стыке современной медицины и математических методов анализа ее данных. Изучение методов, применяемых в медицинской информатике, ориентировано на подготовку студентов к научной и практической работе с современными информационными технологиями в медицине.

Студенты должны:

* ознакомиться с основными целями и принципами медицинской информатики;
* изучить базовые методы представления медицинской информации;
* ознакомиться с основными алгоритмами анализа медицинских изображений;
* ознакомиться с принципами построения и работы медицинских информационных сетей;
* Научиться использовать современные средства обработки медицинских данных и экспорта/импорта данных в формате *DICOM*.

На этом этапе к проведению НИСа также подключаются сотрудники МНУЛ «Интеллектуальные системы и структурный анализ», которые на примере работ стажеров-исследователей и их научных руководителей демонстрируют, каким образом знания в области анализа и обработки медицинских данных могут применяться на практике, например, на примере сотрудничества с *ФНКЦ ДГОИ им. Рогачева*, с которым НИУ ВШЭ проводит совместные исследования по выявлению методик лечения и прогнозирования развития онкологических заболеваний у детей. Данный центр является *первым и крупнейшим в России медицинским учреждением, поддерживающим международные стандарты хранения и обработки медицинской информации*, которая затем используется в международных исследованиях методик и способов лечения онкологических заболеваний.

Студенты изучают методы выделения групп пациентов для улучшений методик лечения относительно показателя выживаемости в группе, а также разрабатывают новые методы обработки и хранения медицинских снимков, полученных при помощи МРТ и ЭЭГ, под руководством профессора Гарвардской медицинской школы О. Пьяных.В заключение, студенты знакомятся с новой для них методикой неинвазивной нейровизуализации – и впервые в рамках курса проводят полевые эксперименты со сбором реальных электроэнцефалографических данных под руководством старшего научного сотрудника центра по нейрокогнитивным исследованиям, А. Осадчего. Методики неинвазивной нейровизуализации на основе ЭЭГ и МЭГ в сочетании с алгоритмами решения обратной задачи обеспечивает беспрецедентное пространственно-временное разрешение, позволяющее детально исследовать быстропротекающие процессы в головном мозге, лежащее в основе экспериментально-контролируемой когнитивной деятельности.

В рамках занятий освещается обзор методов функциональной диагностики мозга (ПЭТ, фЯМР, МРТ, КТ, ЭМЭГ, ЭЭГ) и в рамках открытых лекций НИУ ВШЭ демонстрируют результаты машинного обучения на основе анализа данных, полученных неинвазивным методом (без нарушения целостности тканей) считыванием активности коры головного мозга. Одним из возможных результатов предлагается реализовать набор текста или управление движением в компьютерной игре **силой мысли**.

Программа семинара включает следующие компоненты, связанные с поддержкой написания исследовательских работ:

1. Углубление понимания научного подхода и методологии науки.
2. Подготовка к осознанному выбору темы будущей выпускной квалификационной работы (ВКР). Сюда относится также обзор основных направлений научной деятельности подразделений НИУ ВШЭ, имеющих отношение к реализации научно-исследовательской составляющей концепции ОП «Прикладная математика и информатика».
3. Основы теории чисел и её приложения.
4. Знакомство с основными технологиями и инструментами организации индивидуальной и коллективной научно-исследовательской работы.
5. Подготовка доклада на научной конференции и написание тезисов доклада.

Задача третьего года обучения – обеспечить продолжение профессиональной ориентации студента, снабдить студента необходимым минимумом знаний и компетенций для участия в научных конференциях и прикладных проектах.

## Учебные задачи дисциплины

Данный курс должен помочь сформировать базовые навыки подготовки к выступлению с докладом на конференции и развить навыки научной деятельности.

В результате изучения дисциплины «Научно-исследовательский семинар «Анализ и майнинг данных» студенты должны:

1. понимать принципы научного подхода и методологию научного поиска;
2. уметь адекватно анализировать и оценивать научную деятельность;
3. уметь оформить научный текст и отчет в системе верстки LaTeX, оформить список библиографии в соответствии с ГОСТ;
4. уметь подготовить выступление с использованием средств коллективной работы с текстом и подготовки презентации, и защитить реферат по методам работы с открытыми данными в виде краткого доклада и научной дискуссии.
5. знать стандарт DICOM, способы извлечения и хранения информации в международных стандартах в медицинской информатике;
6. знать основные методы при сжатии изображений и методы кодирования для сжатия с потерями или для безопасной передачи данных;
7. знать основные методы функциональной диагностики мозга (ФДМ): ПЭТ, фЯМР, КТ, ЭМЭГ
8. уметь решать прямую и обратную задачи ЭЭГ и МЭГ;
9. уметь измерить параметры электрического кортикального поля на основе измерений с диполей;
10. владеть навыками прогнозирования решений, принимаемых человеческим мозгом, на основе регистрации и обработки сигналов коры головного мозга.

## Развиваемые компетенции

Дисциплина «Научно-исследовательский семинар «Анализ и майнинг данных» формирует следующие компетенции (классификация 2014 года).

1. ОНК-1: способность к анализу и синтезу на основе системного подхода.
2. ОНК-2: способность перейти от проблемной ситуации к проблемам, задачам и лежащим в их основе противоречиям.
3. ОНК-3: способность использовать методы критического анализа, развития научных теорий, опровержения и фальсификации, оценить качество исследований в некоторой предметной области.
4. ОНК-5: готовность выявить естественно-научную сущность проблем, возникающих в ходе профессиональной деятельности, привлечь их для решения соответствующий физико-математический аппарат.
5. ОНК-6: способность приобретать новые знания с использованием научной методологии и современных образовательных и информационных технологий.
6. ИК-4: способность аналитически работать с информацией из различных источников, включая глобальных компьютерных сетях.
7. СЛК-10: способность использовать в научной и познавательной деятельности, а также в социальной сфере профессиональные навыки работы с информационными и компьютерными технологиями.
8. ПК-3: способность в составе научно-исследовательского и производственного коллектива решать задачи профессиональной деятельности в соответствии с профилем подготовки, общаться с экспертами в других предметных областях.
9. ПК-5: способность осуществлять целенаправленный многокритериальный поиск информации о новейших научных и технологических достижениях в сети Интернет и из других источников.
10. ПК-6: способность собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным, профессиональным, социальным и этическим проблемам.
11. ПК-8: способность решать задачи производственной и технологической деятельности на профессиональном уровне, включая разработку математических моделей, алгоритмических и программных решений.
12. ПК-9: способность применять в профессиональной деятельности современные языки программирования и языки баз данных, операционные системы, электронные библиотеки и пакеты программ, сетевые технологии и т.п.
    * ПК-9-НИС1: способность применять системы управления библиографией.
    * ПК-9-НИС2: способность применять специализированные программные средства подготовки презентаций.
    * ПК-9-НИС3: способность применять специализированные программные средства подготовки научных публикаций.
13. ПК-11: способность составлять и контролировать план выполняемой работы, планировать необходимые для выполнения работы ресурсы, оценивать результаты собственной работы.

# Тематический план курса «Научный семинар (4 курс бакалавриата)»

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№** | **Название темы занятия** | **Всего часов по дисциплине** | **Аудиторные часы** | **Самосто-ятельная работа** |
| **Семинар** |
| 1 | Методология науки. | 6 | 2 | 4 |
| 2 | Методология выбора, подготовки и защиты КР. | 6 | 2 | 4 |
| 3 | Технические и программные средства подготовки КР. | 8 | 2 | 6 |
| 4 | Ведение научной дискуссии и участие в научных мероприятиях. Методология написания тезисов и подготовки выступления на научной конференции. | 6 | 2 | 4 |
| 5 | Введение в медицинскую информатику | 6 | 2 | 4 |
| 6 | Работа с медицинскими изображениями | 6 | 2 | 4 |
| 7 | Автоматизированная диагностика | 8 | 2 | 6 |
| 8 | Безопасность передачи и хранения данных | 6 | 2 | 4 |
| 9 | Методы неинвазивного нейрокартирования | 6 | 2 | 4 |
| 10 | Принципы и устройства для регистрации ЭЭГ и МЭГ. Уравнения Максвелла | 6 | 2 | 4 |
| 11 | Токовый диполь, прямая задача ЭЭГ и МЭГ | 6 | 2 | 4 |
| 12 | Уравнение наблюдения ЭЭГ и МЭГ | 6 | 2 | 4 |
| 13 | Классификация методов решения обратной задачи ЭЭГ и МЭГ | 6 | 2 | 4 |
| 14 | Метод минимальной нормы. Интерпретация в Байесовской парадигме | 6 | 2 | 4 |
| 15 | Улучшения ММН: Loreta, GALA | 8 | 2 | 6 |
| 16 | Параметрические методы решения обратной задачи. Подгонка дипольной модели | 6 | 2 | 4 |
| 17 | Метод неградиентной оптимизации Нелдера-Мида | 6 | 2 | 4 |
| 18 | Решение задачи оценки активации нейрональной популяции по неинвазивным ЭЭГ измерениям при помощи метода формирователя луча | 6 | 2 | 4 |
|  | **Итого** | **72** | **26** | **46** |

# Учебно-методическое и информационное обеспечение дисциплины

## Список литературы

### Базовый учебник

Ридер, составленный по следующим источникам.

1. С.М. Львовский. Набор и верстка в системе LaTeX (<http://www.mccme.ru/free-books/llang/newllang.pdf>)
2. Оформление библиографии (<http://www.academics.hse.ru/bibliography>)
3. Pianykh O. [*Digital Image Quality in Medicine*](https://publications.hse.ru/view/98946833). NY : Springer Verlag, 2013.
4. Oglevee C., Pianykh O. [Losing Images in Digital Radiology: More than You Think](https://publications.hse.ru/view/146991555) // *Journal of Digital Imaging*. 2014. P. 1-8.
5. Zubarev I., Shestakova A., Klucharev V., Ossadtchi A. MEG study of social conformity

<http://publications.hse.ru/chapters/110990446>

### Основная литература

1. Жилин Д. М. Теория систем: опыт построения курса. – КомКнига, 2006. – 184 с.
2. Поппер К. Предположения и опровержения. Рост научного знания. – АСТ, 2004. – 640 с.
3. Лакатос И. Избранные произведения по философии и методологии науки. – Академический проект, 2008. – 480 с.
4. Кун Т. Структура научных революций. – АСТ, 2009. – 320 с.
5. Новиков А.М., Новиков Д.А. Методология научного исследования. – М.: Либроком, 2010. – 280 с.
6. Положение о курсовой работе/курсовом проекте студентов, обучающихся по программам подготовки бакалавров и специалистов, в Национальном исследовательском университете «Высшая школа экономики» (<http://www.hse.ru/docs/28971882.html>)
7. R.S.Ledley, L.B.Lusted, "Reasoning Foundation s of Medical Diagnosis”, Science, July 3 1959
8. R.S.Ledley, "Digital Electronic Computers in Biomedical Science", Science, November 6 1959
9. M.F. Collen, “Origins of medical informatics”, West. J. Med. 145 (1986) 778–785
10. D.F. Sittig, J.S. Ash, R.S.Ledley, “The Story Behind the Development of the First Whole-body Computerized Tomography Scanner as Told by Robert S. Ledley”, Journal of the American Medical Informatics Association Volume 13 Number 5 Sep / Oct 2006
11. M.J. Schuemie, J.L. Talmon, P.W. Moorman, J.A. Kors, Mapping the domain of medical informatics, Methods Inf. Med. 48 (2009) 76–83
12. R. Haux, Medical informatics: Past, present, future, international journal of medical informatics 79 (2010) 599–610
13. Zubarev I., Ossadtchi A., Klucharev V., Shestakova A. MEG signature of social conformity: evidence from evoked and induced responses / Working papers by Центр Нейроэкономики и когнититвных исследований. Series. 1, 1. 2014.
14. A. Ossadtchi, P. Pronko, M. Pflieger, T. Stroganova, Mutual information spectrum – a new tool for detection of event related components in spatial decompositions and its application to M1 cerebral zone localization, Frontiers in Human Neuroscience
15. D. Thyerlei, A. Ossadtchi, T. Maleeva, A.N. Mamelak and W.W. Sutherling, Using intracranial depth electrode stimulation as a reference source for reconstruction from simultaneous scalp-EEG. NeuroImage 2003;
16. Sanei S., Chambers J.A. Eeg Signal Processing. – Wiley, 2007. – 290 p.
17. Neuroscience Online: An Electronic Textbook for the Neurosciences (<http://neuroscience.uth.tmc.edu>)

### Дополнительная литература

1. Новиков А.М., Новиков Д.А. Методология научного исследования. – М.: Либроком, 2010. – 280 с.
2. Кузнецов И. Н. Научное исследование: Методика проведения и оформление. – 2-е изд. – М.: Дашков и Ко, 2006. – 460 с.
3. Заметки к лекции Б.Г. Миркина «Структура научного доклада и основы научного обсуждения».
4. Колесникова Н.И. От конспекта к диссертации: Учебное пособие по развитию навыков письменной речи. – М.: Флинта: Наука, 2009. – 288 с.
5. Center for Open Science (<http://centerforopenscience.org>)
6. The OpenScience Project (<http://www.openscience.org>)
7. J.P. DeShazo, D.L. LaVallie, F.M.Wolf, Publication trends in the medical informatics literature: 20 years of "Medical Informatics" in MeSH, BioMedCentral, Januaru 21, 2009.
8. Bansard JY, Rebholz-Schuhmann D, Cameron G, Clark D, van Mulligen E, Beltrame E, Barbolla E, Martin-Sanchez Fdel H, Milanesi L, Tollis I, van der Lei J, Coatrieux JL, Medical informatics and bioinformatics: a bibliometric study, IEEE Trans Inf Technol Biomed. 2007 May;11(3):237-43. Review
9. Ossadtchi A. Connectivity measures applied to human brain electrophysiological data // Journal of Neuroscience Methods, 207(1), 2012, pp. 1-16.
10. The OpenScience Project (<http://www.openscience.org>)
11. Frictionless Open Data (<http://data.okfn.org>)
12. The Open Source Data Science Masters – Curriculum for Data Science (<http://datasciencemasters.org>)

# Формы контроля и структура итоговой оценки

**Текущий контроль**

В 1-м модуле оценивается реферат по одному из способов хранения (полнота описания – 20%), обработки (сравнение методов и программных продуктов – 20%) и извлечения данных (сравнение с собственной реализацией – 20%). Реферат выполняется в среде коллективной работы ShareLaTeX и сопровождается презентацией в пакете Beamer. 40% оценки за реферат составляет оценка студентами выступления участника и его уровень ведения научной дискуссиии.

Во 2-м модуле выполняются 4 проекта, каждый из которых оценивается преподавателем индивидуально (25% каждый). Проект по изучению стандарта DICOM оценивается на основе полноты разбора представления данных. Проекты по обработке 2D и 3D изображений и автоматической диагностике пациентов оцениваются по количественным метрикам соотношения сжатие объема/потеря качества изображения и точности кластеризации патологий – оценка выставляется по системе grade-by-queue как соответствующие проценты от результатов в группе, но не ниже требуемых в условии значений (для оценки «отлично»). Проекты по цифровым меткам и безопасности проверяются экспериментально на основе результатов изучения работы [Mapping Digital Radiology Adoption and Security Worldwide](http://www.ajronline.org/doi/abs/10.2214/AJR.15.15283) и апробации системы защиты данных, или исследования технологии цифровых меток на примере открытых данных.

В 3-4 модулях каждое ДЗ (предполагается 3-4 дз с равными коэффициентами и вложенные по содержанию решаемых заданий) проверяется «старшими» кураторами по проектам от департамента психологии и департамента анализа данных и искусственного интеллекта. Результаты методов прогнозирования моделей сравниваются внутри каждого из проектов и шкалируются внутри группы.

**Итоговый контроль** – 1 экзамен виде публичного представления проекта (в конце четвертого модуля); в случае невыполнения проекта (и низкой накопленной пропускавших занятия студентов) проводится устный экзамен по темам 3-4 модулей, содержащим основные научные направления семинара. Итоговый проект на публичное представление отбирает специальное жюри. Каждый из проектов оценивается научными сотрудниками центра НКИ и МНУЛ ИССА вместе с преподавателями семинара.

Итоговая оценка складывается из следующих элементов:

*Отекущий* = 0,2*·О1 модуль +* 0,4*·О2 модуль +*  0,4·*О3-4 модули*

*Одисциплина =* 0,2·*Оэкзамен +* 0,8·*Отекущий*

Д/з сдаётся не позднее установленной преподавателем даты (зависит от графика учебного процесса и объявляется при выдаче задания). Д/з защищается путём демонстрации отчёта и ответов на вопросы преподавателя по отчёту.

На пересдаче неудовлетворительной оценки за д/з студенту предоставляется возможность получить не более 3 дополнительных баллов для компенсации оценки за текущий контроль. Дата пересдачи определяется преподавателем (зависит от графика учебного процесса и объявляется после проверки отчётов по д/з). Пересдача д/з допускается только один раз.

Посещение менее 50% обязательных научных мероприятий (семинаров, указанных преподавателем) ведет к выставлению текущей оценки не выше, чем оценка «удовлетворительно». Пропуск обязательных научных мероприятий необходимо компенсировать сдачей индивидуального отчета по пропущенным темам.

### Таблица соответствия оценок по десятибалльной и системе зачет/незачет

|  |  |
| --- | --- |
| **Оценка по 10-балльной шкале** | **Оценка по 5-балльной шкале** |
| 1 | незачет |
| 2 |
| 3 |
| 4 | зачет |
| 5 |
| 6 |
| 7 |
| 8 |
| 9 |
| 10 |

### Таблица соответствия оценок по десятибалльной и пятибалльной системе

|  |  |
| --- | --- |
| **По десятибалльной шкале** | **По пятибалльной системе** |
| 1 – неудовлетворительно  2 – очень плохо  3 – плохо | неудовлетворительно – 2 |
| 4 – удовлетворительно  5 – весьма удовлетворительно | удовлетворительно – 3 |
| 6 – хорошо  7 – очень хорошо | хорошо – 4 |
| 8 – почти отлично  9 – отлично  10 – блестяще | отлично – 5 |

# Программа дисциплины

## Тема 1. Методология науки

История развития прикладной математики и информатики. Научный метод. Гипотезы, теории и их фальсифицируемость. Эксперименты и их воспроизводимость. Методология научных исследований.

Основные направления развития компьютерных наук. Научные группы на факультете компьютерных наук НИУ ВШЭ.

Введение в современные технологии поддержки научно-исследовательской деятельности.

### Основная литература

1. Жилин Д. М. Теория систем: опыт построения курса. – КомКнига, 2006. – 184 с.
2. Поппер К. Предположения и опровержения. Рост научного знания. – АСТ, 2004. – 640 с.
3. Лакатос И. Избранные произведения по философии и методологии науки. – Академический проект, 2008. – 480 с.
4. Кун Т. Структура научных революций. – АСТ, 2009. – 320 с.

### Дополнительная литература

1. Новиков А.М., Новиков Д.А. Методология научного исследования. – М.: Либроком, 2010. – 280 с.
2. The OpenScience Project (http://www.openscience.org)

## Тема 2. Методология выбора, подготовки и защиты КР

Ознакомление с регламентом работы над междисциплинарной курсовой работой (КР) и Положением о курсовой работе/курсовом проекте студентов, обучающихся по программам подготовки бакалавров и специалистов, в Национальном исследовательском университете «Высшая школа экономики».

Методические рекомендации по написанию курсовой работы. Требования к КР. Процесс подготовки КР. Особенности защиты КР на 2 курсе бакалавриата и в дальнейшем. От КР к выпускной квалификационной работе (ВКР).

Дополнительное знакомство с основными научными направлениями в рамках концепции ОП Прикладной математики и информатики. Посещение обязательных научных мероприятий факультета/департамента.

### Основная литература

1. Сайт факультета компьютерных наук (<http://cs.hse.ru>)
2. Положение о курсовой работе/курсовом проекте студентов, обучающихся по программам подготовки бакалавров и специалистов, в Национальном исследовательском университете «Высшая школа экономики» (<http://www.hse.ru/docs/28971882.html>)

### Дополнительная литература

1. Кузнецов И. Н. Научное исследование: Методика проведения и оформление. – 2-е изд. – М.: Дашков и Ко, 2006. – 460 с.
2. Center for Open Science (<http://centerforopenscience.org>)

## Тема 3. Технические и программные средства подготовки КР

Введение в систему подготовки текстов *LaTeX*. Основные дистрибутивы. Базовые возможности, работа со стилями, оформление математических формул. Вставка иллюстраций.

Работа с основными облачными сервисами *Microsoft* и *Google*, системами *Microsoft OneNote* и *EverNote* и другими коллаборативными технологиями.

Подготовка списка литературы с использованием библиографических менеджеров *BibTeX*, *JabRef*, *Mendeley*, *Zotero* и др.

Основы работы в *Microsoft PowerPoint* и пакете *Biber* для *LaTeX*. Подготовка 3-х слайдовой промежуточной презентации по итогам выбора тем и постановки задачи КР.

Оформление аннотации, общей концепции, плана и введения, списка литературы с помощью изученных программных средств.

### Основная литература

1. С.М. Львовский. Набор и верстка в системе LaTeX (<http://www.mccme.ru/free-books/llang/newllang.pdf>).
2. Авторские материалы по подготовке презентаций.
3. Оформление библиографии (http://www.academics.hse.ru/bibliography)

### Дополнительная литература

1. Колесникова Н.И. От конспекта к диссертации: Учебное пособие по развитию навыков письменной речи. – М.: Флинта: Наука, 2009. – 288 с.

## Тема 4. Ведение научной дискуссии и участие в научных мероприятиях

Основы ведения научной дискуссии. Семинары и конференции.

Ключевые моменты подготовки выступления на научном мероприятии. Основные правила поведения на защите КР, вынесение результатов на защиту КР. Дистанционная защита КР.

Приобретение опыта выступлений и ведения дискуссий.

Рецензирование и исправление замечаний рецензента в онлайн-сервисе *ShareLaTeX* ([www.sharelatex.com](http://www.sharelatex.com)).

### Основная литература

1. Лакатос И. Избранные произведения по философии и методологии науки. – Академический проект, 2008. – 480 с.

### Дополнительная литература

1. Заметки к лекции Б.Г. Миркина «Структура научного доклада и основы научного обсуждения».
2. Колесникова Н.И. От конспекта к диссертации: Учебное пособие по развитию навыков письменной речи. – М.: Флинта: Наука, 2009. – 288 с.
3. Сайт *EasyChair* (<http://www.easychair.org>)

## Тема 5. Анализ медицинских данных.

I. Что такое Медицинская Информатика и зачем она нужна? Образцы DICOM-снимков (в DICOM-формате).

II. Создание хорошего снимка. Сжатие с потерями. Прототип программы фильтрации, и снимки для неею

III. Computer-Aided Diagnostics. Автоматизированная диагностика на основе анализа изображений.

IV. Медицина на расстоянии

V.\* Задачи оптимизации в медицине

Все материалы доступны и обновляются на сайте курса на странице автора.

### Базовый учебник

Ридер, составленный по следующим источникам.

1. Pianykh O. [*Digital Image Quality in Medicine*](https://publications.hse.ru/view/98946833). NY : Springer Verlag, 2013.
2. Oglevee C., Pianykh O. [Losing Images in Digital Radiology: More than You Think](https://publications.hse.ru/view/146991555) // *Journal of Digital Imaging*. 2014. P. 1-8.
3. Stites, S., Pianykh, O.S. How Secure Is Your Radiology Department? Mapping Digital Radiology Adoption and Security Worldwide. American Journal of Roentgenology: 1-8. 10.2214/AJR.15.15283

### Основная литература

1. R.S.Ledley, L.B.Lusted, "Reasoning Foundation s of Medical Diagnosis”, Science, July 3 1959
2. R.S.Ledley, "Digital Electronic Computers in Biomedical Science", Science, November 6 1959
3. M.F. Collen, “Origins of medical informatics”, West. J. Med. 145 (1986) 778–785
4. D.F. Sittig, J.S. Ash, R.S.Ledley, “The Story Behind the Development of the First Whole-body Computerized Tomography Scanner as Told by Robert S. Ledley”, Journal of the American Medical Informatics Association Volume 13 Number 5 Sep / Oct 2006
5. M.J. Schuemie, J.L. Talmon, P.W. Moorman, J.A. Kors, Mapping the domain of medical informatics, Methods Inf. Med. 48 (2009) 76–83
6. R. Haux, Medical informatics: Past, present, future, international journal of medical informatics 79 (2010) 599–610

### Дополнительная литература

1. J.P. DeShazo, D.L. LaVallie, F.M.Wolf, Publication trends in the medical informatics literature: 20 years of "Medical Informatics" in MeSH, BioMedCentral, Januaru 21, 2009.
2. Bansard JY, Rebholz-Schuhmann D, Cameron G, Clark D, van Mulligen E, Beltrame E, Barbolla E, Martin-Sanchez Fdel H, Milanesi L, Tollis I, van der Lei J, Coatrieux JL, Medical informatics and bioinformatics: a bibliometric study, IEEE Trans Inf Technol Biomed. 2007 May;11(3):237-43. Review

## Тема 6. Реализация методов обработки сигналов в задаче обратной связи

Оригинальный цикл лекций о направлении, находящемся на острие прогресса – *Neuroscience*.

### Занятие 1. Обзор методов функциональной диагностики мозга (ФДМ)

Использование ФДМ в медицинской практике и исследованиях.

Понятие пространственного и временного разрешения.

Описание и сравнение методов ФДМ (ПЭТ, фЯМР, КТ, ЭМЭГ).

### Занятие 2. Происхождение сигналов, регистрируемых методом электро и магнитоэнцефалографии

Природа электрических сигналов, регистрируемых ЭМЭГ.

Оборудование необходимое для проведения ЭМЭГ экспериментов.

Процесс измерения.

Сопоставление с анатомической информацией (ЯМР).

Основные экспериментальные парадигмы, в т.ч. парадигмы реального времени (нейрообратная связь и мозг-компьютерный интерфейс).

Наиболее частые ошибки, проблемы и их решения.

### Занятие 3. Прямая задача в ЭМЭГ, методы моделирования ЭМЖЭГ данных

Аппроксимация, методы расчета прямой модели.

Модель наблюдения ЭМЭГ сигнала.

### Занятие 4. Пространственная и временная фильтрация, удаление артефактов

Спектральное представление сигналов (Преобразование Фурье, Wavelet преобразование).

Временная фильтрация (КИХ, БИХ фильтры, фильтры в частотной и wavelet областях).

Пространственная фильтрация (метод главных компонент, метод независимых переменных).

### Занятие 5. Обратная задача ЭМЭГ

Модель распределённого источника.

Байесовская формулировка обратной задачи.

Различные алгоритмы как следствие различных априорных предположений.

Описание различных алгоритмов метода наименьшей нормы (MN, WMN, FOCUSS, LORETA, GALA).

### Занятие 6. Параметрический подход к решению обратной задачи

Дипольная модель.

Модель движущегося диполя.

Метод наименьших квадратов для подгонки модели.

Методы, использующие понятие подпространства сигнала (MUSIC, RAP-MUSIC).

### Занятие 7. Парадигмы реального времени и методы обработки сигнала

Мозг-компьютерный интерфейс – классификация парадигм.

Методы обработки сигнала для извлечения признаков.

Методы классификации.

### Основная литература

1. Sanei S., Chambers J.A. Eeg Signal Processing. – Wiley, 2007. – 290 p.
2. Neuroscience Online: An Electronic Textbook for the Neurosciences (<http://neuroscience.uth.tmc.edu>)

### Дополнительная литература

1. Ossadtchi A. Connectivity measures applied to human brain electrophysiological data // Journal of Neuroscience Methods, 207(1), 2012, pp. 1-16.

# Тематика заданий по формам текущего контроля

### Оформление тезисов доклада на научной конференции

Тезисы объёмом не более 2 страниц, посвящённые решению конкретной задачи. Структура тезисов следующая.

1. Актуальность и мотивация.
2. Цель работы, постановка задачи и её обоснование .
3. Основные методы решения и обоснование проектных решений.
4. Главное отличие от ранее предлагавшихся решений.
5. Результаты (включая теоретическую значимость, практическую полезность и апробацию).
6. Варианты развития работы и планы на будущее.
7. Уточнение содержания доклада на конференции.

### Оформление списка использованных источников

При этом требуется:

1. знание ГОСТ 7.1-2003 и структуры библиографического описания;
2. знание ГОСТ 7.05-2008 и правильное формирование затекстовых ссылок;
3. стилевых пакетов ugost2003, ugost2008s.

### Оформление презентации доклада

Презентация на основе тезисов доклада на конференции.

### Выступление на научном семинаре

Краткий рассказ по выбранной тематике исследований.

**Вопросы к теоретическому экзамену** (для тех, кто не сделал проект).

**Методология операций с данными**

**Данные как объект исследования**

**Что такое данные**

1. Информация и данные
2. Данные и алгоритмы
3. Объём данных
4. Задачи обработки данных
   1. Место анализа данных - в **переработке** данных
5. Телекоммуникации
6. Коллаборативные технологии

**Характеристики данных**

1. Тип
2. Формат
3. Объём
4. Источник
5. Изменение

**История вопроса**

От вычислительных и логистических задач WW2 до Big Data и Data Science.

**Форматы данных**

1. Standards, specifications and RFCs
   1. Main Data Types
      1. Common
         1. A Tutorial on Data Representation Integers, Floating-point Numbers, and Characters ([https://www3.ntu.edu.sg/home/ehchua/programming/java/DataRepresentation.html](https://mail2.hse.ru/owa/redir.aspx?SURL=7viFPcVJoD7qezFBPZCHGmhNWlyxySt77i07rSbwinOJahEJxlHTCGgAdAB0AHAAcwA6AC8ALwB3AHcAdwAzAC4AbgB0AHUALgBlAGQAdQAuAHMAZwAvAGgAbwBtAGUALwBlAGgAYwBoAHUAYQAvAHAAcgBvAGcAcgBhAG0AbQBpAG4AZwAvAGoAYQB2AGEALwBEAGEAdABhAFIAZQBwAHIAZQBzAGUAbgB0AGEAdABpAG8AbgAuAGgAdABtAGwA&URL=https%3a%2f%2fwww3.ntu.edu.sg%2fhome%2fehchua%2fprogramming%2fjava%2fDataRepresentation.html))
      2. Strings
         1. UNICODE ([http://www.unicode.org](https://mail2.hse.ru/owa/redir.aspx?SURL=bBnwdWWS-41Kp6URdTzyYPiclZ5rkPLl2FIpty_cjlWJahEJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgB1AG4AaQBjAG8AZABlAC4AbwByAGcALwA.&URL=http%3a%2f%2fwww.unicode.org%2f))
            1. ICU - International Components for Unicode ([http://site.icu-project.org](https://mail2.hse.ru/owa/redir.aspx?SURL=0RobhiUZh-6pdsO5dp4EGu-FI1KNnECNadbJA0EV1L-JahEJxlHTCGgAdAB0AHAAOgAvAC8AcwBpAHQAZQAuAGkAYwB1AC0AcAByAG8AagBlAGMAdAAuAG8AcgBnAC8A&URL=http%3a%2f%2fsite.icu-project.org%2f))
         2. ISO/IEC 10646:2014 Information technology - Universal Coded Character Set (UCS) ([http://www.iso.org/iso/home/store/catalogue\_ics/catalogue\_detail\_ics.htm?csnumber=63182](https://mail2.hse.ru/owa/redir.aspx?SURL=NYQefCBUkxyd0xquNfNk_xZ0ITbb7RDM9R3Go5GFM3SJahEJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgBpAHMAbwAuAG8AcgBnAC8AaQBzAG8ALwBoAG8AbQBlAC8AcwB0AG8AcgBlAC8AYwBhAHQAYQBsAG8AZwB1AGUAXwBpAGMAcwAvAGMAYQB0AGEAbABvAGcAdQBlAF8AZABlAHQAYQBpAGwAXwBpAGMAcwAuAGgAdABtAD8AYwBzAG4AdQBtAGIAZQByAD0ANgAzADEAOAAyAA..&URL=http%3a%2f%2fwww.iso.org%2fiso%2fhome%2fstore%2fcatalogue_ics%2fcatalogue_detail_ics.htm%3fcsnumber%3d63182))
            1. Specifies three encoding forms of the UCS: UTF-8, UTF-16, and UTF-32
      3. Date/Time
         1. ISO Time - ISO 8601:1988 "Data elements and interchange formats - Information interchange - Representation of dates and times" ([http://www.iso.org/iso/ru/home/standards/iso8601.htm](https://mail2.hse.ru/owa/redir.aspx?SURL=l1XMESx2G6nql9cUeNwfKWqqouL2rsPJVUhiOl6yyRyJahEJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgBpAHMAbwAuAG8AcgBnAC8AaQBzAG8ALwByAHUALwBoAG8AbQBlAC8AcwB0AGEAbgBkAGEAcgBkAHMALwBpAHMAbwA4ADYAMAAxAC4AaAB0AG0A&URL=http%3a%2f%2fwww.iso.org%2fiso%2fru%2fhome%2fstandards%2fiso8601.htm))
            1. Date and Time Formats ([http://www.w3.org/TR/NOTE-datetime](https://mail2.hse.ru/owa/redir.aspx?SURL=uDFmcX3IXHYB4bo2XTCHU0mM7kukraIhNntuLq7FXIOJahEJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgB3ADMALgBvAHIAZwAvAFQAUgAvAE4ATwBUAEUALQBkAGEAdABlAHQAaQBtAGUA&URL=http%3a%2f%2fwww.w3.org%2fTR%2fNOTE-datetime))
      4. Numbers
         1. Floating Point Number - IEEE 754 ([http://grouper.ieee.org/groups/754](https://mail2.hse.ru/owa/redir.aspx?SURL=ZoC6uk6CDBazX2QKDljZp9KddvvklmgnT0wod_SI3CaJahEJxlHTCGgAdAB0AHAAOgAvAC8AZwByAG8AdQBwAGUAcgAuAGkAZQBlAGUALgBvAHIAZwAvAGcAcgBvAHUAcABzAC8ANwA1ADQA&URL=http%3a%2f%2fgrouper.ieee.org%2fgroups%2f754))
   2. Data Serialization
      1. XML
         1. XML 1.0 - W3C Recommendation 2008-11-26 ([http://www.w3.org/TR/REC-xml](https://mail2.hse.ru/owa/redir.aspx?SURL=sIUTBRPhae7_SrBwEOaxxy8RtvC841VNczVnr67oRySJahEJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgB3ADMALgBvAHIAZwAvAFQAUgAvAFIARQBDAC0AeABtAGwA&URL=http%3a%2f%2fwww.w3.org%2fTR%2fREC-xml))
         2. W3Schools XML Tutorial ([http://www.w3schools.com/xml](https://mail2.hse.ru/owa/redir.aspx?SURL=TbeKsaqle10-eshzwqOMku-jvDtyLJ6dTEyG-LRyDvaJahEJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgB3ADMAcwBjAGgAbwBvAGwAcwAuAGMAbwBtAC8AeABtAGwA&URL=http%3a%2f%2fwww.w3schools.com%2fxml))
         3. The Annotated XML 1.0 Specification 1998-02-10 ([http://www.xml.com/axml/testaxml.htm](https://mail2.hse.ru/owa/redir.aspx?SURL=ASNMa8PeiJGpjX2kqFii3JcErijSbkaRqdP5jiP8TQGJahEJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgB4AG0AbAAuAGMAbwBtAC8AYQB4AG0AbAAvAHQAZQBzAHQAYQB4AG0AbAAuAGgAdABtAA..&URL=http%3a%2f%2fwww.xml.com%2faxml%2ftestaxml.htm))
         4. Namespaces in XML 1.0 (Third Edition) W3C Recommendation 2009-12-8 ([http://www.w3.org/TR/REC-xml-names](https://mail2.hse.ru/owa/redir.aspx?SURL=WAgqTUFQOYZwhkxIpRsNaBKsuwzTyx0dLjwLI3CAXQGJahEJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgB3ADMALgBvAHIAZwAvAFQAUgAvAFIARQBDAC0AeABtAGwALQBuAGEAbQBlAHMA&URL=http%3a%2f%2fwww.w3.org%2fTR%2fREC-xml-names))
         5. Document Object Model (DOM) Level 2 Core Specification (Version 1.0) W3C Recommendation 2000-11-13 ([http://www.w3.org/TR/DOM-Level-2-Core](https://mail2.hse.ru/owa/redir.aspx?SURL=f9ZMNDYkLHckh92UoQ0r25n1vNc51jpzBujlIXuC22SJahEJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgB3ADMALgBvAHIAZwAvAFQAUgAvAEQATwBNAC0ATABlAHYAZQBsAC0AMgAtAEMAbwByAGUA&URL=http%3a%2f%2fwww.w3.org%2fTR%2fDOM-Level-2-Core))
         6. Microsoft XML Standards Reference ([https://msdn.microsoft.com/en-us/library/ms256177(v=vs.110).aspx](https://mail2.hse.ru/owa/redir.aspx?SURL=wdR7tLHiF7PaB1GoS9i3PJkPiIFjFRv7CPBFEAfAK7eJahEJxlHTCGgAdAB0AHAAcwA6AC8ALwBtAHMAZABuAC4AbQBpAGMAcgBvAHMAbwBmAHQALgBjAG8AbQAvAGUAbgAtAHUAcwAvAGwAaQBiAHIAYQByAHkALwBtAHMAMgA1ADYAMQA3ADcAKAB2AD0AdgBzAC4AMQAxADAAKQAuAGEAcwBwAHgA&URL=https%3a%2f%2fmsdn.microsoft.com%2fen-us%2flibrary%2fms256177%28v%3dvs.110%29.aspx))

1. CSV
   1. [CSV] CSV – Comma Separated Values ([http://data.okfn.org/doc/csv](https://mail2.hse.ru/owa/redir.aspx?SURL=5itQ-dO5yah9idT1eYWfxWBgjAbXMvfxve4TFzSTCZ6JahEJxlHTCGgAdAB0AHAAOgAvAC8AZABhAHQAYQAuAG8AawBmAG4ALgBvAHIAZwAvAGQAbwBjAC8AYwBzAHYA&URL=http%3a%2f%2fdata.okfn.org%2fdoc%2fcsv))
   2. [CSV] RFC 4180 - Common Format and MIME Type for Comma-Separated Values (CSV) Files ([http://tools.ietf.org/html/rfc4180](https://mail2.hse.ru/owa/redir.aspx?SURL=bnY0fUimxmIxF-o-k0N24CIDsiyYPXNYIkoepVZ97rqJahEJxlHTCGgAdAB0AHAAOgAvAC8AdABvAG8AbABzAC4AaQBlAHQAZgAuAG8AcgBnAC8AaAB0AG0AbAAvAHIAZgBjADQAMQA4ADAA&URL=http%3a%2f%2ftools.ietf.org%2fhtml%2frfc4180))
   3. [CSV] RFC 7111 - URI Fragment Identifiers for the text/csv Media Type ([https://tools.ietf.org/html/rfc7111](https://mail2.hse.ru/owa/redir.aspx?SURL=S6XsWlUbt2YQgGrmTYTWFDK_OXZohyWIZJ8yMYVfkXOJahEJxlHTCGgAdAB0AHAAcwA6AC8ALwB0AG8AbwBsAHMALgBpAGUAdABmAC4AbwByAGcALwBoAHQAbQBsAC8AcgBmAGMANwAxADEAMQA.&URL=https%3a%2f%2ftools.ietf.org%2fhtml%2frfc7111))

1. JSON
   1. JSON - ECMA-404 ([http://json.org](https://mail2.hse.ru/owa/redir.aspx?SURL=ET_fI7CoHjLTNg58H9tiPLRp5la547QXjNh8-ZUkRSOJahEJxlHTCGgAdAB0AHAAOgAvAC8AagBzAG8AbgAuAG8AcgBnAC8A&URL=http%3a%2f%2fjson.org%2f) , [http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf](https://mail2.hse.ru/owa/redir.aspx?SURL=5Q-nyZ1JxFlI5JeD3V3xef8HoWNulqfX2nlr78jyC5CJahEJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgBlAGMAbQBhAC0AaQBuAHQAZQByAG4AYQB0AGkAbwBuAGEAbAAuAG8AcgBnAC8AcAB1AGIAbABpAGMAYQB0AGkAbwBuAHMALwBmAGkAbABlAHMALwBFAEMATQBBAC0AUwBUAC8ARQBDAE0AQQAtADQAMAA0AC4AcABkAGYA&URL=http%3a%2f%2fwww.ecma-international.org%2fpublications%2ffiles%2fECMA-ST%2fECMA-404.pdf))
   2. [JSON] JSON-LD 1.0 - A JSON-based Serialization for Linked Data - W3C Recommendation 16 January 2014 ([http://www.w3.org/TR/json-ld/](https://mail2.hse.ru/owa/redir.aspx?SURL=gJ4rUVO2jfnu917fTF2HXjJcj9Ns9cP0NDTk6Ks8DSCJahEJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgB3ADMALgBvAHIAZwAvAFQAUgAvAGoAcwBvAG4ALQBsAGQALwA.&URL=http%3a%2f%2fwww.w3.org%2fTR%2fjson-ld%2f))
   3. [JSON] BSON ([http://bsonspec.org/spec.html](https://mail2.hse.ru/owa/redir.aspx?SURL=eBWvLcvOgvD954K856NCnB6-QFQY88Y2eDl46q3Hd_SJahEJxlHTCGgAdAB0AHAAOgAvAC8AYgBzAG8AbgBzAHAAZQBjAC4AbwByAGcALwBzAHAAZQBjAC4AaAB0AG0AbAA.&URL=http%3a%2f%2fbsonspec.org%2fspec.html), [http://bsonspec.org](https://mail2.hse.ru/owa/redir.aspx?SURL=sxTg5jPTaQzBzRsYZO1a7f9S9aDBKFC7X066OjsWTRaJahEJxlHTCGgAdAB0AHAAOgAvAC8AYgBzAG8AbgBzAHAAZQBjAC4AbwByAGcALwA.&URL=http%3a%2f%2fbsonspec.org%2f))
   4. [JSON] UBJSON ( [http://en.wikipedia.org/wiki/UBJSON](https://mail2.hse.ru/owa/redir.aspx?SURL=5wG6pIh0AvgOM69FSskUXUl8puUsMUSUnT7z2bxulkCJahEJxlHTCGgAdAB0AHAAOgAvAC8AZQBuAC4AdwBpAGsAaQBwAGUAZABpAGEALgBvAHIAZwAvAHcAaQBrAGkALwBVAEIASgBTAE8ATgA.&URL=http%3a%2f%2fen.wikipedia.org%2fwiki%2fUBJSON))
   5. [JSON] Hjson, the Human JSON ([http://hjson.org](https://mail2.hse.ru/owa/redir.aspx?SURL=C6DP7DM_o2QjpRXAyIA8HuiCvXMGGwQA0QR0XH9BdvKJahEJxlHTCGgAdAB0AHAAOgAvAC8AaABqAHMAbwBuAC4AbwByAGcALwA.&URL=http%3a%2f%2fhjson.org%2f))
   6. [JSON] JSON5 ([http://json5.org](https://mail2.hse.ru/owa/redir.aspx?SURL=EnS82_I-2LmvGCe9vS8xWgSbFyWegVyD-u-U33v0KJKJahEJxlHTCGgAdAB0AHAAOgAvAC8AagBzAG8AbgA1AC4AbwByAGcALwA.&URL=http%3a%2f%2fjson5.org%2f))
   7. [JSON] JSON Schema ([http://json-schema.org](https://mail2.hse.ru/owa/redir.aspx?SURL=Hx_dANuDe-dsazL2nf1Ofd_QKIsMjRA9fqa4LeR1k4CJahEJxlHTCGgAdAB0AHAAOgAvAC8AagBzAG8AbgAtAHMAYwBoAGUAbQBhAC4AbwByAGcALwA.&URL=http%3a%2f%2fjson-schema.org%2f))

1. ASN.1 ([http://www.itu.int/en/ITU-T/asn1/Pages/asn1\_project.aspx](https://mail2.hse.ru/owa/redir.aspx?SURL=oUAsyb36flvw7-nsqcUUk6UE6LVl2niRJySBv0m-r66JahEJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgBpAHQAdQAuAGkAbgB0AC8AZQBuAC8ASQBUAFUALQBUAC8AYQBzAG4AMQAvAFAAYQBnAGUAcwAvAGEAcwBuADEAXwBwAHIAbwBqAGUAYwB0AC4AYQBzAHAAeAA.&URL=http%3a%2f%2fwww.itu.int%2fen%2fITU-T%2fasn1%2fPages%2fasn1_project.aspx))
2. MessagePack ([http://msgpack.org](https://mail2.hse.ru/owa/redir.aspx?SURL=ViB9rn1_f1zWkbbidSN5EVC5Um15N0qCZ61x-PYoHaPkyxMJxlHTCGgAdAB0AHAAOgAvAC8AbQBzAGcAcABhAGMAawAuAG8AcgBnAC8A&URL=http%3a%2f%2fmsgpack.org%2f))
3. Configuration files
   1. INI Files
      1. Windows API GetPrivateProfileString function ([https://msdn.microsoft.com/en-us/library/windows/desktop/ms724353.aspx](https://mail2.hse.ru/owa/redir.aspx?SURL=-eKbJ12HhXoS0uMwh8MdFSRLHzf3CLDbrLOHFBx--IrkyxMJxlHTCGgAdAB0AHAAcwA6AC8ALwBtAHMAZABuAC4AbQBpAGMAcgBvAHMAbwBmAHQALgBjAG8AbQAvAGUAbgAtAHUAcwAvAGwAaQBiAHIAYQByAHkALwB3AGkAbgBkAG8AdwBzAC8AZABlAHMAawB0AG8AcAAvAG0AcwA3ADIANAAzADUAMwAuAGEAcwBwAHgA&URL=https%3a%2f%2fmsdn.microsoft.com%2fen-us%2flibrary%2fwindows%2fdesktop%2fms724353.aspx))
      2. Cloanto Implementation of INI File Format ([http://cloanto.com/specs/ini/](https://mail2.hse.ru/owa/redir.aspx?SURL=-u28aEa-J9NUbxZyKw1GfjrvcRALznYGMDY3Jd-Q1nnkyxMJxlHTCGgAdAB0AHAAOgAvAC8AYwBsAG8AYQBuAHQAbwAuAGMAbwBtAC8AcwBwAGUAYwBzAC8AaQBuAGkALwA.&URL=http%3a%2f%2fcloanto.com%2fspecs%2fini%2f))
      3. minIni ([https://github.com/compuphase/minIni](https://mail2.hse.ru/owa/redir.aspx?SURL=_GtmMx_k73UULUX-ZG24vPwhS3KUv5mk8SKORklipOHkyxMJxlHTCGgAdAB0AHAAcwA6AC8ALwBnAGkAdABoAHUAYgAuAGMAbwBtAC8AYwBvAG0AcAB1AHAAaABhAHMAZQAvAG0AaQBuAEkAbgBpAA..&URL=https%3a%2f%2fgithub.com%2fcompuphase%2fminIni))
   2. TOML - Tom's Obvious, Minimal Language by Tom Preston-Werner ([https://github.com/toml-lang/toml](https://mail2.hse.ru/owa/redir.aspx?SURL=eEX-_WPN_FLLdxqs7naCdtKEXhevDBZW-XSkaJt3TCXkyxMJxlHTCGgAdAB0AHAAcwA6AC8ALwBnAGkAdABoAHUAYgAuAGMAbwBtAC8AdABvAG0AbAAtAGwAYQBuAGcALwB0AG8AbQBsAA..&URL=https%3a%2f%2fgithub.com%2ftoml-lang%2ftoml))
4. [Additional info]
   1. How to Populate a Property Tree ([http://www.boost.org/doc/libs/1\_42\_0/doc/html/boost\_propertytree/parsers.html](https://mail2.hse.ru/owa/redir.aspx?SURL=mm4uUXE48vNr7k5anU07FQ_oYa-kQ1Vzpr6bEKb9eO7kyxMJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgBiAG8AbwBzAHQALgBvAHIAZwAvAGQAbwBjAC8AbABpAGIAcwAvADEAXwA0ADIAXwAwAC8AZABvAGMALwBoAHQAbQBsAC8AYgBvAG8AcwB0AF8AcAByAG8AcABlAHIAdAB5AHQAcgBlAGUALwBwAGEAcgBzAGUAcgBzAC4AaAB0AG0AbAA.&URL=http%3a%2f%2fwww.boost.org%2fdoc%2flibs%2f1_42_0%2fdoc%2fhtml%2fboost_propertytree%2fparsers.html))
5. Other formats
   1. Microformats 2 ([http://microformats.org/wiki/microformats2](https://mail2.hse.ru/owa/redir.aspx?SURL=w9IH8fiXbe1wXD01uKHeCaTohB0A00u_GhUEjlKRWavkyxMJxlHTCGgAdAB0AHAAOgAvAC8AbQBpAGMAcgBvAGYAbwByAG0AYQB0AHMALgBvAHIAZwAvAHcAaQBrAGkALwBtAGkAYwByAG8AZgBvAHIAbQBhAHQAcwAyAA..&URL=http%3a%2f%2fmicroformats.org%2fwiki%2fmicroformats2))
6. Guidelines and tutorials
   1. Web-services data serialization…
   2. MSDN - Introducing XML Serialization ([https://msdn.microsoft.com/en-us/library/182eeyhh(v=vs.110).aspx](https://mail2.hse.ru/owa/redir.aspx?SURL=oHYLHsLsuXg0UPtosBYmyugafzCiayvnZwjZMousc9jkyxMJxlHTCGgAdAB0AHAAcwA6AC8ALwBtAHMAZABuAC4AbQBpAGMAcgBvAHMAbwBmAHQALgBjAG8AbQAvAGUAbgAtAHUAcwAvAGwAaQBiAHIAYQByAHkALwAxADgAMgBlAGUAeQBoAGgAKAB2AD0AdgBzAC4AMQAxADAAKQAuAGEAcwBwAHgA&URL=https%3a%2f%2fmsdn.microsoft.com%2fen-us%2flibrary%2f182eeyhh%28v%3dvs.110%29.aspx))

**Предобработка данных [ETL]**

Extract Transform Load

**Свойства данных, затрудняющие их анализ**

1. **Гетерогенность источников.** Необходимость работы с данными из различных источников с различными форматами и API.
2. **Big Data.** Необходимость работы с данными больших объемов, что затрудняет задачи этапа ETL для последующей обработки и визуализации.
3. **Итеративность.** Итеративность большинства методологий анализа и необходимость хранения промежуточных результатов анализа на каждом этапе.
4. **Синонимия.** Существование множества названий для одного и того явления или понятия.
5. **Омонимия.** Существование различных понятий с одним именем.
6. **Строгие ограничения на время для анализа данных.** Анализ данных представляет ценность при принятии решений. В связи с этим как правило существует ограничения на время для анализа данных.

**Открытые данные [open data]**

**History of term**

1. History of the **Open Definition** started in 2005 ([http://opendefinition.org/history](https://mail2.hse.ru/owa/redir.aspx?SURL=n9y_popkFZq0_soC_i2LHTxjRSh_fYn1xSQMaseZCD7kyxMJxlHTCGgAdAB0AHAAOgAvAC8AbwBwAGUAbgBkAGUAZgBpAG4AaQB0AGkAbwBuAC4AbwByAGcALwBoAGkAcwB0AG8AcgB5AA..&URL=http%3a%2f%2fopendefinition.org%2fhistory))
2. Memorandum for the Heads of Executive Departments and Agencies by Barak Obama in 2008: **Transparency and Open Government**,  ([https://www.whitehouse.gov/the\_press\_office/TransparencyandOpenGovernment](https://mail2.hse.ru/owa/redir.aspx?SURL=GCUKsM0rwAbr5KDZMXmNK7O1ZiLzzoZxuxFiWEeqP1HkyxMJxlHTCGgAdAB0AHAAcwA6AC8ALwB3AHcAdwAuAHcAaABpAHQAZQBoAG8AdQBzAGUALgBnAG8AdgAvAHQAaABlAF8AcAByAGUAcwBzAF8AbwBmAGYAaQBjAGUALwBUAHIAYQBuAHMAcABhAHIAZQBuAGMAeQBhAG4AZABPAHAAZQBuAEcAbwB2AGUAcgBuAG0AZQBuAHQA&URL=https%3a%2f%2fwww.whitehouse.gov%2fthe_press_office%2fTransparencyandOpenGovernment))
3. On the stage of the 2009 TED conference, Sir Tim Berners-Lee, the inventor of the hypertext links and ultimately, of the Web, made the following rallying cry (for the supporters of data Web) or war cry (for data holders): “We want raw data, now!” (video: Tim Berners-Lee, TED, 2009 - [http://www.dailymotion.com/video/x8omty\_tim-berners-lee-the-next-web-of-ope\_tech#.UVWqxhnGrJw](https://mail2.hse.ru/owa/redir.aspx?SURL=vZnce3jzLd7hdNZCfUusUD1TddCvlOyI6526-BeLJurkyxMJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgBkAGEAaQBsAHkAbQBvAHQAaQBvAG4ALgBjAG8AbQAvAHYAaQBkAGUAbwAvAHgAOABvAG0AdAB5AF8AdABpAG0ALQBiAGUAcgBuAGUAcgBzAC0AbABlAGUALQB0AGgAZQAtAG4AZQB4AHQALQB3AGUAYgAtAG8AZgAtAG8AcABlAF8AdABlAGMAaAAjAC4AVQBWAFcAcQB4AGgAbgBHAHIASgB3AA..&URL=http%3a%2f%2fwww.dailymotion.com%2fvideo%2fx8omty_tim-berners-lee-the-next-web-of-ope_tech%23.UVWqxhnGrJw))
4. Early projects
   1. The **Human Genome Project** in the late 1990s and early 2000s which was fully open ([http://www.sanger.ac.uk/about/history/hgp](https://mail2.hse.ru/owa/redir.aspx?SURL=N9D11ioxQRX1O-ia_bF628HmbqKos3lXVEpB3DX6DPDkyxMJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgBzAGEAbgBnAGUAcgAuAGEAYwAuAHUAawAvAGEAYgBvAHUAdAAvAGgAaQBzAHQAbwByAHkALwBoAGcAcAA.&URL=http%3a%2f%2fwww.sanger.ac.uk%2fabout%2fhistory%2fhgp))
   2. In December 2002, **Creative Commons** released its first set of copyright licenses ([https://creativecommons.org/about/history](https://mail2.hse.ru/owa/redir.aspx?SURL=gzwKoj6Gm6DbZFbm7LvudHWDosTCqg7DaDhdOUDt_3fkyxMJxlHTCGgAdAB0AHAAcwA6AC8ALwBjAHIAZQBhAHQAaQB2AGUAYwBvAG0AbQBvAG4AcwAuAG8AcgBnAC8AYQBiAG8AdQB0AC8AaABpAHMAdABvAHIAeQA.&URL=https%3a%2f%2fcreativecommons.org%2fabout%2fhistory))
   3. **Open Street Map**, started in late 2004, perhaps the most successful community Open Data project to date ([**http://openstreetmap.org**](https://mail2.hse.ru/owa/redir.aspx?SURL=XtqMqbEhUitfcH6upsLDyEKl2lY28gGI4IJ1l2vHHQbkyxMJxlHTCGgAdAB0AHAAOgAvAC8AbwBwAGUAbgBzAHQAcgBlAGUAdABtAGEAcAAuAG8AcgBnAC8A&URL=http%3a%2f%2fopenstreetmap.org%2f))
   4. The **Open Knowledge Forums** in London in 2005-2006 organized by the Open Knowledge Foundation which covered topics like open geodata and open civic information ([http://okfn.org/events/okforums](https://mail2.hse.ru/owa/redir.aspx?SURL=x7srPEsJANkt3sU48ZrmyvIADA9NZe3viiVPQZFPhUPkyxMJxlHTCGgAdAB0AHAAOgAvAC8AbwBrAGYAbgAuAG8AcgBnAC8AZQB2AGUAbgB0AHMALwBvAGsAZgBvAHIAdQBtAHMA&URL=http%3a%2f%2fokfn.org%2fevents%2fokforums))
5. Reviews
   1. Simon Chignard, A brief history of Open Data ([http://www.paristechreview.com/2013/03/29/brief-history-open-data](https://mail2.hse.ru/owa/redir.aspx?SURL=SadjAlIzTq4kHSIV0ryZm2T90tJhVe_PvsO92SUJFU3kyxMJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgBwAGEAcgBpAHMAdABlAGMAaAByAGUAdgBpAGUAdwAuAGMAbwBtAC8AMgAwADEAMwAvADAAMwAvADIAOQAvAGIAcgBpAGUAZgAtAGgAaQBzAHQAbwByAHkALQBvAHAAZQBuAC0AZABhAHQAYQA.&URL=http%3a%2f%2fwww.paristechreview.com%2f2013%2f03%2f29%2fbrief-history-open-data))
   2. Defining “Open” by DAVID on NOVEMBER 16, 2009 ([http://opencontent.org/blog/archives/1123](https://mail2.hse.ru/owa/redir.aspx?SURL=jmFB9345CVA0y2Iu1OKhl9xJkZZ_aDyTRQKdxny8GerkyxMJxlHTCGgAdAB0AHAAOgAvAC8AbwBwAGUAbgBjAG8AbgB0AGUAbgB0AC4AbwByAGcALwBiAGwAbwBnAC8AYQByAGMAaABpAHYAZQBzAC8AMQAxADIAMwA.&URL=http%3a%2f%2fopencontent.org%2fblog%2farchives%2f1123))
   3. Yong Shi, Big Data History, Current Status, and Challenges going Forward ([https://www.nae.edu/Publications/Bridge/128772/129211.aspx](https://mail2.hse.ru/owa/redir.aspx?SURL=-ogHsMKeYiUEya_lhnvMFZjC83OMEV0kjBjR7iYkjFrkyxMJxlHTCGgAdAB0AHAAcwA6AC8ALwB3AHcAdwAuAG4AYQBlAC4AZQBkAHUALwBQAHUAYgBsAGkAYwBhAHQAaQBvAG4AcwAvAEIAcgBpAGQAZwBlAC8AMQAyADgANwA3ADIALwAxADIAOQAyADEAMQAuAGEAcwBwAHgA&URL=https%3a%2f%2fwww.nae.edu%2fPublications%2fBridge%2f128772%2f129211.aspx))

Основные принципы использования открытых данных, принятые правительством США

Открытые данные должны удовлетворять требованиям:

1. **Публичности** (Public)

Публикация открытых данных должна подразумевать согласие с законами о приватности, конфиденциальности и безопасности, а также соответствие другим правомерным ограничениям. [Open Government Directive, 8.12.2009, ([http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda\_2010/m10-06.pdf](https://mail2.hse.ru/owa/redir.aspx?SURL=QT-okYYKKImDFaWhrQqPfvWMdIG9RlO0RE2jJSUDVgXkyxMJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgB3AGgAaQB0AGUAaABvAHUAcwBlAC4AZwBvAHYALwBzAGkAdABlAHMALwBkAGUAZgBhAHUAbAB0AC8AZgBpAGwAZQBzAC8AbwBtAGIALwBhAHMAcwBlAHQAcwAvAG0AZQBtAG8AcgBhAG4AZABhAF8AMgAwADEAMAAvAG0AMQAwAC0AMAA2AC4AcABkAGYA&URL=http%3a%2f%2fwww.whitehouse.gov%2fsites%2fdefault%2ffiles%2fomb%2fassets%2fmemoranda_2010%2fm10-06.pdf))]

1. **Доступности** (Accessible)

Открытые данные должны храниться в удобных открытых форматах, предполагающих простоту модификации (внесения изменений), индексации, поиска и скачивания (получения через глобальную сеть Интернет). Используемые форматы данных должны быть машинно-читаемыми, то есть позволяющими проводить полностью автоматическую обработку. Структура данных должна покрывать как можно более широкий спектр задач и возможных пользователей, что может потребовать предоставления открытых данных в различных форматах. Форматы в любом случае не должны накладывать каких-либо дополнительных ограничений на обработку данных, то есть быть открытыми, публично-доступными, не защищёнными патентами.

1. **Адекватности описания (полноты метаданных)** (Described)

Открытые данные должны быть полностью описаны таким образом, чтобы пользователь имел достаточно полную информацию о том, как обрабатывать эти данные, об их сильных и слабых сторонах, ограничениях, требованиях безопасности. Это подразумевает наличие метаданных, описывающих тезаурус, структуру и отдельные поля данных, а также, по возможности, цели сбора данных, параметры генеральной совокупности и выборки, методы сбора данных.

1. **Разрешённости для повторного использования** (Reusable)

Открытые данные должны предоставляться в соответствии с открытыми лицензиями, которые не ограничивают последующее использование данных.

1. **Полноты** (Complete)

Открытые данные должны быть доступны в своей исходной форме с максимально возможным уровнем детализации, практически достижимым и не противоречащим требованиям законов и прочим ограничениям. Производные и агрегированные открытые данные должны явно ссылаться на исходные данных.

1. **Актуальности** (Timely)

Открытые данные должны обновляться с частотой, необходимой для сохранения непротиворечивости, согласованности и ценности данных. Частота обновления данных должна учитывать потребности в данных ключевой аудитории.

1. **Управляемости после публикации** (Managed Post-Release)

Контактное лицо или организация должны предоставлять помощь по использованию данных и отвечать на жалобы о несоблюдении вышеприведённых свойств. A point of contact must be designated to assist with data use and to respond to complaints about adherence to these open data requirements.

**4R Framework**

Источник <[http://opencontent.org/blog/archives/1123](https://mail2.hse.ru/owa/redir.aspx?SURL=jmFB9345CVA0y2Iu1OKhl9xJkZZ_aDyTRQKdxny8GerkyxMJxlHTCGgAdAB0AHAAOgAvAC8AbwBwAGUAbgBjAG8AbgB0AGUAbgB0AC4AbwByAGcALwBiAGwAbwBnAC8AYQByAGMAaABpAHYAZQBzAC8AMQAxADIAMwA.&URL=http%3a%2f%2fopencontent.org%2fblog%2farchives%2f1123)>

The 4Rs Framework describes the four most important rights:

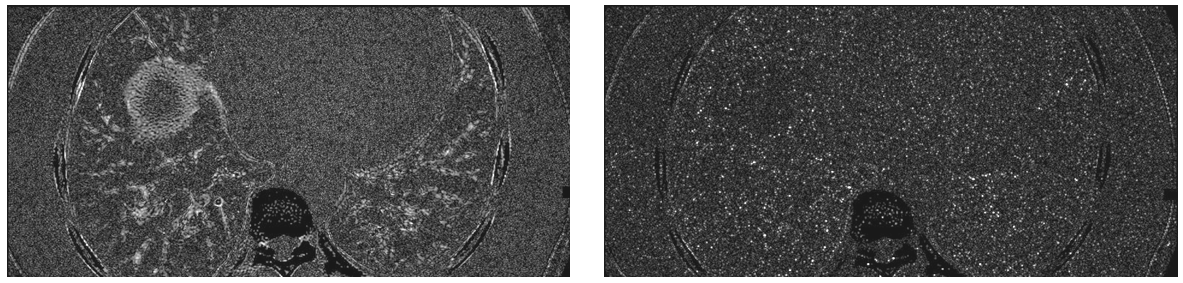
1. **Reuse** – the right to reuse the content in its unaltered / verbatim form.
2. **Revise** – the right to adapt, adjust, modify, or alter the content itself.
3. **Remix** – the right to combine the original or revised content with other content to create something new.
4. **Redistribute** – the right to make and share copies of the original content, your revisions, or your remixes with others.

**Примеры открытых данных**

1. Open data 500 ([http://www.opendata500.com](https://mail2.hse.ru/owa/redir.aspx?SURL=N0KXXDejm07P0LyZFcWxs4UCGcAPCOyrbIDTsS42OxvkyxMJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgBvAHAAZQBuAGQAYQB0AGEANQAwADAALgBjAG8AbQAvAA..&URL=http%3a%2f%2fwww.opendata500.com%2f))
2. Ships ([https://pendientedemigracion.ucm.es/info/cliwoc/](https://mail2.hse.ru/owa/redir.aspx?SURL=_gsvCG5sBgJcsHcQ7XdFtJzXrNofMVbvu0HxmTLtOjvkyxMJxlHTCGgAdAB0AHAAcwA6AC8ALwBwAGUAbgBkAGkAZQBuAHQAZQBkAGUAbQBpAGcAcgBhAGMAaQBvAG4ALgB1AGMAbQAuAGUAcwAvAGkAbgBmAG8ALwBjAGwAaQB3AG8AYwAvAA..&URL=https%3a%2f%2fpendientedemigracion.ucm.es%2finfo%2fcliwoc%2f))
3. Live Ships Map - AIS - Vessel Traffic and Positions - AIS Marine Traffic ([http://www.marinetraffic.com/](https://mail2.hse.ru/owa/redir.aspx?SURL=-3MDo9-P8s9R1QLIjyszq0eWDNUluAdYqpO3QVt8u-3kyxMJxlHTCGgAdAB0AHAAOgAvAC8AdwB3AHcALgBtAGEAcgBpAG4AZQB0AHIAYQBmAGYAaQBjAC4AYwBvAG0ALwA.&URL=http%3a%2f%2fwww.marinetraffic.com%2f))
4. Flight Radar 24 ([https://www.flightradar24.com](https://mail2.hse.ru/owa/redir.aspx?SURL=aQCyJTnGb3UrZ43QcKqaszIaorimaUPsWmTbC_qIbUjkyxMJxlHTCGgAdAB0AHAAcwA6AC8ALwB3AHcAdwAuAGYAbABpAGcAaAB0AHIAYQBkAGEAcgAyADQALgBjAG8AbQAvAA..&URL=https%3a%2f%2fwww.flightradar24.com%2f))
5. Cities
   1. OPEN DATA of Moscow City Government ([http://data.mos.ru](https://mail2.hse.ru/owa/redir.aspx?SURL=5KDBLilvQWf2HTekeFyUyx1Lme0kRizZ2BHKYr5lslPkyxMJxlHTCGgAdAB0AHAAOgAvAC8AZABhAHQAYQAuAG0AbwBzAC4AcgB1AC8A&URL=http%3a%2f%2fdata.mos.ru%2f))
   2. NYC Open Data ([https://nycopendata.socrata.com/dashboard](https://mail2.hse.ru/owa/redir.aspx?SURL=NhTBtOqbSJQa9SUyXD_pElyi3HX3E9s9FcMumEQPxx7kyxMJxlHTCGgAdAB0AHAAcwA6AC8ALwBuAHkAYwBvAHAAZQBuAGQAYQB0AGEALgBzAG8AYwByAGEAdABhAC4AYwBvAG0ALwBkAGEAcwBoAGIAbwBhAHIAZAA.&URL=https%3a%2f%2fnycopendata.socrata.com%2fdashboard))
   3. London DataStore ([http://data.london.gov.uk](https://mail2.hse.ru/owa/redir.aspx?SURL=7E9T121Romu7PPeS2whMW6LMqiJzIngTqkfI9vRcpxbkyxMJxlHTCGgAdAB0AHAAOgAvAC8AZABhAHQAYQAuAGwAbwBuAGQAbwBuAC4AZwBvAHYALgB1AGsALwA.&URL=http%3a%2f%2fdata.london.gov.uk%2f))

**Medical Data Analysis**

1. Which of the following text fragments is HL7?
   1. MSH|^~\&|ADT|N|ADT|MEDSC|200601081527
   2. <html><body><h1>Report</h1><p>PatientName.</p></body></html>
   3. <script type="text/javascript">document.write("<p>"+Date()+"</p>");
   4. r = input('Enter radius: '); a = pi\*r^2;
2. Sinogram for a straight line segment is a
   1. Point
   2. Triangle
   3. Square
   4. Segment
3. Two different algorithms, A and B, were used to remove noise from a diagnostic image. The following are the difference images (Original minus Denoised) for each algorithm. Would you recommend A or B?



A

B

1. A
2. B
3. Recall that to regularize an image, we should minimize the following function, with positive *ukn*:

*E( Y ) =* Σ*n* [ *( xn - yn )*2 *+ λ* Σ*k ukn ( yk - yn )*2 ]

If we replace the original image *X* by filtered image *Y* such that E(*Y*) is minimized when *λ* = 0.5, what conclusions can you make about the original and filtered images?

1. One can do tomography without X-Rays:
   1. True
   2. False



1. The following equation defines a bilateral filter:
   1. True
   2. False
2. Which of the following properties is often enforced with noisy image regularization:
3. Sharpness
4. Smoothness
5. Lambdaness
6. Brightness
7. It is critical to migrate to the most recent versions of medical standards right after they are released, because:
8. They provide more functionality
9. They advance clinical workflow
10. This is a law
11. No, it is not critical
12. Temporal data analysis in Radiology becomes possible because
13. Time stamps are stored in DICOM images
14. Time stamps are registered by special time-stamping devices
15. All modalities scan images with the same timing intervals
16. None of the above
17. What are the two principal types of DICOM data encoding?
18. VR and non-VR
19. Private and Standard
20. Implicit and Explicit
21. Compressed and Uncompressed
22. You need to start a teleradiology project, sending CT images to a remote clinic for reports. You measured the network link speed to that clinic; it is 10MB/sec on average. So you consider image compression to send your images faster. The compression software takes 2 seconds to compress (or uncompress) 100 CT images, with average Rcomp=3.2. Is it worth using, and why?  
    CT image size is 0.5 MB.
23. Yes
24. No
25. It is critical to migrate to the most recent versions of medical standards right after they are released, because:
    1. They provide more functionality
    2. They advance clinical workflow
    3. This is a law
    4. No, it is not critical
26. Temporal data analysis in Radiology becomes possible because
    1. Time stamps are stored in DICOM images
    2. Time stamps are registered by special time-stamping devices
    3. All modalities scan images with the same timing intervals
    4. None of the above
27. A CT image with “bits allocated” equal to 16 was compressed so that in the compressed format it requires only 2 bits per pixel. What is the compression ratio?
    1. 1
    2. 2
    3. 3
    4. 4
28. Diagnostic image compression ensures that
    1. The original color palette is preserved
    2. The lossy compression error at each pixel is bounded by a chosen threshold
    3. The compression will use only lossless algorithms
    4. The image size remains unchanged
29. The following formula defines
30. Entropy



1. Standard Gaussian filter
2. LoG filter
3. Denoising filter
4. Which type of texture analysis will most likely rely on PCA?
5. Parametric
6. Non-parametric
7. The following fragment of Matlab code was written to compute bilateral filter for image L at point (i,j):

for ri = -6:6

for rj = -6:6

dL = (L(i, j) - L(i + ri, j + rj))/200; %L1

w = lambda \* exp( dL\*dL ) / (1 + rj \* rj + ri \* ri); %L2

s1 = s1 + w; %L3

s2 = s2 + w \* L(i + ri, j + rj); %L4

end

end

end

Which line contains an error?

1. L1
2. L2
3. L3
4. L4
5. Non-exact conditions in a scheduling program are known as
   1. Hard Constraints
   2. Soft Constraints
   3. Relaxed Constraints
   4. Fixed Constraints
6. What is PHI?
   1. Patient health information
   2. Protected health information
   3. Patient HL7 information
   4. Private health instruction
7. What is HIPAA ?
   1. Health Insurance Privacy and Accountability Act
   2. Health Insurance Privacy Access Act
   3. Health Insurance Portability and Accountability Act
   4. HL7 Insurance Portability and Accountability Act
8. Which of the following is/are considered PHI?
9. Web Universal Resource Locators (URLs)
10. Certificate / license numbers
11. Health plan beneficiary numbers
12. All of the above
13. One needs special DICOM software to find patient name in DICOM files
14. True
15. False
16. Can one restore the original data back from the anonymized?
    1. True
    2. False
17. Image watermarking (as discussed in class) is based on:
18. X-Rays
19. e-Paper
20. Anonymization
21. Steganography
22. Find errors (as many as you see)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Byte#** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| **Decimal** | 16 | 0 | 16 | 0 | P | N | 10 | 0 | S | m | i | t | h | ^ | J | o | e |
| **Binary** | 10 | 00 | 00 | 10 | 50 | 4E | 0A | 00 | 53 | 6D | 69 | 74 | 68 | 5E | 4A | 6F | 65 |
|  | g=0010 | | e=0010 | | VR type | | VR length L=10=0x000A | | VR value=Smith^Joe | | | | | | | | | |

1. Even data length
2. No trailing semicolon
3. Wrong name separator
4. VR tyoe must not be present
5. Watermarking cannot be destroyed by lossy compression
   1. True
   2. False

**Neuroscience**

1. Методы неинвазивного нейрокартирования. Фундаментальные принципы используемых методик, сравнительный анализ пространственного и временного разрешения.

2. Что измеряет ЭЭГ и МЭГ? Принципы и устройства для регистрации ЭЭГ. Принципы и устройства для регистрации МЭГ. Понятие первичного тока и вторичных токов. Отличие ЭЭГ от МЭГ в контексте решения обратной задачи.

3. Уравнения Максвелла, смысл каждого из уравнений, смысл операторов вихря и дивергенции (rot и div), квазистатическая аппроксимация уравнений Максвелла и решение для сферического однослойного проводника (без вывода).

4. Токовый диполь, прямая задача ЭЭГ и МЭГ. Матрица прямой модели. Варианты матрицы для свободной и фиксированной ориентации диполей. Анатомическая подоплека использования фиксированной ориентации. Смысл столбцов и строк матрицы прямой модели.

5. Понятия токового диполя, топографии и временной последовательности активации нейронального источника. Уравнение наблюдения ЭЭГ и МЭГ. Пространственно белый и пространственно коррелированный шум. Модель шума в уравнении наблюдения.

6. Классификация методов решения обратной задачи ЭЭГ и МЭГ. Основные предположения и допущения, используемые в двух основных классах методов.

7. Метод минимальной нормы. Фундаментальные предположения и интерпретация в Байесовской парадигме, как задачи оценивания по принципу максимальной апостериорной вероятности (MAP).

8. Варианты метода Минимальной Нормы (Minimum Norm Estimation [MNE]): Loreta (в деталях), GALA (общие принципы).

9. Параметрические методы решения обратной задачи. Подгонка дипольной модели — две основные метрики, используемые при подгонке дипольной модели. Методика определения числа диполей, необходимых для описания наблюдаемой матрицы ЭЭГ(МЭГ) данных.

10. Метод неградиентной оптимизации Нелдера-Мида и его применение при подгонке диполей. Как используется факт линейной зависимости прямой модели от параметра активации нейрональных источников при подгонке диполей?

11. Решение задачи оценки активации нейрональной популяции по неинвазивным ЭЭГ(МЭГ) измерениям при помощи метода формирователя луча (Beamforming). Сканирующий формирователь луча для решения обратной задачи. Объяснение причин теоретически достижимого супер-разрешения при решении обратной задачи. Сравнение с методом минимальной нормы.

# Вопросы для оценки качества освоения дисциплины

1. Что такое научная теория? Чем она отличается от гипотезы?
2. Что такое методология?
3. Что такое парадигма?
4. Что такое научная группа и научная школа?
5. Почему в методологии науки столь важное значение придается понятию «фальсифицируемость»?
6. Что такое модель, математическая модель, решение модели?
7. Какие приложения математических методов Вы знаете?
8. Какие направления прикладных математических исследований наиболее актуальны в настоящее время?
9. Что такое вейвлет?
10. Объясните отличия преобразования Фурье от вевлет-преобразования.
11. Как регистрируется мозговая активность?
12. Что такое энцефалограмма?
13. Какие методы считаются перспективными для построения мозг-компьютерных интерфейсов?
14. Почему задача дискретного логарифмирования является практически важной?
15. Сформулируйте отличия компьютерных наук от других наук (если эти отличия существуют).
16. Как информационные технологии соотносятся с математикой? Почему прикладных математиков часто «путают» с программистами? Что вкладывается в понятие «программист» в данном случае?
17. Что такое «проблема останова»?
18. Как доказывается корректность алгоритмов?
19. Какие исследования проводятся на ОП «Прикладная математика и информатика»?
20. Какие формы научной коммуникации Вы знаете?
21. Какие типы научных мероприятий Вы знаете?
22. В чём состоят основные особенности научной дискуссии и публичной защиты результатов научно-исследовательской работы?
23. Что такое тезис, антитезис, аргументы и контраргументы?
24. Кратко сформулируйте Ваши научные интересы.
25. Почему Вы выбрали данную тему КР и данного научного руководителя?
26. Технические вопросы по работе в системе подготовки текстов *LaTeX*.
27. Технические вопросы по работе с приложением *PowerPoint* и пакетом Biber создания презентаций в LaTeX– эффективное создание презентаций в соответствии с методологией.

# Методические указания студентам

## О посещении обязательных научных мероприятий

При **невозможности посетить** какое-либо из обязательных научных мероприятий по уважительной причине следует обратиться к ведущему секции и/или преподавателю, отвечающему за дисциплину.

**Пропуск** обязательных научных мероприятий **без уважительной причины** необходимо компенсировать в виде **отработки** (письменного сообщения и, при необходимости, доклада) по пропущенным темам и сдать лично преподавателю, отвечающему за дисциплину.

Все материалы лекций дублируются на сайт курса. Некоторые из предлагаемых для изучения материалов (в особенности, читаемые иностранным профессором), написаны на английском языке.

## Регистрация в различных Интернет-сервисах

Для выполнения учебных заданий необходимо зарегистрироваться в нескольких Интернет-сервисах. Минимальный набор приведён ниже.

1. *SharLaTex* (<https://www.sharelatex.com>)
2. *EasyChair* (<https://www.easychair.org>)

## Критерии оценки презентации

1. Использование шаблона оформления (отсутствие немотивированных отступлений от шаблона).
2. Контрастная цветовая схема (оптимум – черные символы на белом фоне).
3. Первый слайд: название, авторы [,научный руководитель].
4. Последний слайд: обобщающая информация, контакты, благодарности.
5. Название, соответствующее докладу, а не просто повторение темы общего проекта (общая тема может быть указана дополнительно).
6. Грамматика (отсутствие точек в конце заголовков; отсутствие точек и точек с запятой в конце элементов списков (вне абзаца!); полные абзацы текста вне списков оформлены по стандартным правилам; и др.
7. Заголовки слайдов (согласованные заголовки в разделе, состоящем из нескольких слайдов).
8. Подписи слайдов, колонтитулы (оптимум – копирайт, автор, дата презентации) и номера слайдов.
9. Слайд со структурой презентации, если длительность выступления больше 15 мин.
10. Подписи рисунков и таблиц, не занимающих полный слайд.
11. Правильное оформление формул, включая полные легенды.
12. Правильное оформление цитирования (только непосредственное цитирование – без ссылок на отдельный слайд).

# Topics for research projects in Harvard Hospital

* Diagnostic image compression
* Predictive models for patient flow and resource allocation
* Study of human errors in medical data
* 3D medical image segmentation
* Codeless teleradiology gateway. This project is illustrated in the appendix below
* Noise removal in CT images to reduce patient radiation exposure
* Basic Computer-Aided Diagnostics for liver tumors

Автор программы: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/ Незнанов А.А. /

Автор программы: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/ Пьяных О.С. /

Автор программы: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/ Осадчий А.Е. /