

Софийски университет „Св. Климент Охридски“

Факултет по математика и информатика

Курсов проект

на тема: „Онтология на заболяванията по МКБ-10“

Студент: Мартин Попов

Факултетен номер: 6MI3400517

Курс: „Представяне и моделиране на знания“

Учебна година: 2025/26

Преподаватели: проф. Мария Нишева, ас. Мелания Бербатова

Съдържание

1 Въведение и цели.....	2
2 Набор на данни.....	2
3 Моделиране.....	7
4 Основни класове и йерархия.....	8
5 Обектни свойства (релации)	8
6 Свойства по типа на данни.....	9
7 Индивиди и извеждане (Reasoning).....	9
8 Компетентностни въпроси и заявки.....	9
9 Визуализация (WebOWL)	10
10 Бъдещо развитие	11
11 Съдържание, генерирано с изкуствен интелект	11
12 Използвана литература.....	11

1 Въведение и цели

Целта на онтологията е формално и структурирано представяне на знания, описани в Международната класификация на болестите (ICD-10) и допълващи медицински източници. Медицинската информация е формално структурирана, така че да позволява търсене и анализ извън стандартното текстово представяне на МКБ-10.

Онтологията има следните цели:

- консистентно моделиране на юерархията на МКБ-10 (глави, блокове, категории и подкатегории)
- формално описание на заболяванията и връзките им със симптоми, органови системи и етиология
- автоматично класифициране и reasoning с OWL 2 DL reasoner, включително извеждане на принадлежност към дефинирани групи (напр. инфекциозни, респираторни)
- контрол на коректността на данните
- поддръжка на DL заявки за аналитично и диагностично търсене в медицински контекст

Едно съществуващо решение е:

<https://bioportal.bioontology.org/ontologies/ICD10?p=summary>

Това решение е доста постно и не включва свойства като симптоми, етиологични фактори или органова система.

Има също и DiseaseOntology - <https://disease-ontology.org/>, което външност се доближава доса до идеята, но нейната идея е по-скоро да свързва заболяванията с външни стандарти, като МКБ-10, докато нашата онтология се фокусира главно върху този стандарт.

2 Набор на данни

Има няколко варианта за това:

- използване на МКБ-10 кодове (клинична модификация)

Този вариант съдържа ~72 хиляди различни кода. Има набор от данни с всички заболявания [2], но е твърде подробен и разширен, което доведе до по-лоши резултати. Затова този вариант беше оставен като евентуално бъдещо разширение

- Използване на приложно-програмен интерфейс за МКБ-10, предоставен от CZO [3]

Събирането на данните протече в следните стъпки:

- 1) Намиране на последната актуална версия (2019)
Заявка към <http://id.who.int/icd/release/10> [1]

```
{  
    "@context": "http://id.who.int/icd-contexts/contextForMultiVersion.json",  
    "@id": "http://id.who.int/icd/release/10",  
    "title": {  
        "@language": "en",  
        "@value": "International Statistical Classification of Diseases and  
Related Health Problems 10th Revision (ICD-10)"  
    },  
    "latestRelease": "http://id.who.int/icd/release/10/2019",  
    "release": [  
        "http://id.who.int/icd/release/10/2019",  
        "http://id.who.int/icd/release/10/2016",  
        "http://id.who.int/icd/release/10/2010",  
        "http://id.who.int/icd/release/10/2008"  
    ]  
}
```

- 2) Запитване към последната актуална версия, запознаване със структурата на резултатите от приложно-програмния интерфейс (всеки резултат има поле "child" – масив от други валидни линкове, към което може да се правят заявки)

Заявка към: <http://id.who.int/icd/release/10/2019>

```
{  
    "@context": "http://id.who.int/icd-contexts/contextForTopLevel.json",  
    "@id": "http://id.who.int/icd/release/10/2019",  
    "title": {  
        "@language": "en",  
        "@value": "International Statistical Classification of Diseases and  
Related Health Problems 10th Revision (ICD-10) Version for 2019"  
    },  
    "releaseDate": "2020-02-01",  
    "child": [  
        "http://id.who.int/icd/release/10/2019/I",  
        ...  
        "http://id.who.int/icd/release/10/2019/XXII"  
    ],  
    "browserUrl": "http://apps.who.int/classifications/icd10/browse/2019/en"  
}
```

3) Заявки към децата докато полето не стане празно или просто го няма
Заявка към: <http://id.who.int/icd/release/10/2019/I>

```
{  
    "@id": "http://id.who.int/icd/release/10/2019/I",  
    "parent": [  
        "http://id.who.int/icd/release/10/2019"  
    ],  
    "child": [  
        "http://id.who.int/icd/release/10/2019/A00-A09",  
        ...  
        "http://id.who.int/icd/release/10/2019/B99-B99"  
    ],  
    "code": "I",  
    "title": {  
        "@value": "Certain infectious and parasitic diseases"  
    },  
    "inclusion": [  
        ...  
    ],  
    "exclusion": [  
        ...  
    ],  
    "classKind": "chapter"  
}
```

Заявка към: <http://id.who.int/icd/release/10/2019/A00>

```
{  
    "@id": "http://id.who.int/icd/release/10/2019/A00",  
    "parent": [  
        "http://id.who.int/icd/release/10/2019/A00-A09"  
    ],  
    "child": [  
        "http://id.who.int/icd/release/10/2019/A00.0",  
        "http://id.who.int/icd/release/10/2019/A00.1",  
        "http://id.who.int/icd/release/10/2019/A00.9"  
    ],  
    "browserUrl":  
    "http://apps.who.int/classifications/icd10/browse/2019/en#/A00",  
    "code": "A00",  
    "title": {  
        "@value": "Cholera"  
    },  
    "classKind": "category"  
}
```

Заявка към: <http://id.who.int/icd/release/10/2019/A00.0>

```
{  
    "@id": "http://id.who.int/icd/release/10/2019/A00.0",  
    "parent": [  
        "http://id.who.int/icd/release/10/2019/A00"  
    ],  
    "browserUrl":  
    "http://apps.who.int/classifications/icd10/browse/2019/en#/A00.0",  
    "code": "A00.0",  
    "title": {  
        "@value": "Cholera due to Vibrio cholerae 01, biovar cholerae"  
    },  
    "inclusion": [  
        {  
            "label": {  
                "@value": "Classical cholera"  
            }  
        }  
    ],  
    "classKind": "category"  
}
```

4) Агрегиране на информацията за да получим накрая пълни данни за един код

Имаме 4 вида ресурси – глава (*chapter*), блок (*block*), категория (*category*), подкатегория (*subcategory*).

Всеки има различно име. Например:

Chapter I - Certain infectious and parasitic diseases

Block A00-A09 - Intestinal infectious diseases

Category A00 – Cholera

Subcategory A00.0 - Cholera due to Vibrio cholerae 01, biovar cholerae

Понеже търсенето е отгоре-надолу (*chapter* → *block* → *category* → *subcategory*) за най-ниското ниво нямаме информация в резултата към коя глава или кой блок принадлежи. Затова, по време на scraping-а агрегираме данните, така че накрая да получим пълна информация за даден код по МКБ-10.

Набяването на симптомите за различни заболявания се оказа по-сложна задача, понеже доста заболявания имат сходни симптоми, а и никъде особено няма набор от данни, който директно да свързва заболяване със симптоми. Бяха намерени няколко корпуса с данни, които имаха симптоми за заболявания, но бяха представени в различни видове:

- име на заболяване + симптоми [4], [5], [6] – беше нужно запитване към частта на проекта, където по описание/име на заболяване вземаме код по МКБ-10
- код на заболяване по *Disease Ontology (DOID code)* [7] + симптоми [8] – включващо създаване на mapping между *DOID* и МКБ-10 и навръзване на резултатите

Агрегираните данни излеждат по следния начин:

```
{  
    "icd10code": "F07.9",  
    "type": "subcategory",  
    "title": "Unspecified organic personality and behavioural disorder due to  
brain disease, damage and dysfunction",  
    "chapter": "Mental and behavioural disorders",  
    "chapterCode": "V",  
    "block": "Organic, including symptomatic, mental disorders",  
    "blockCode": "F00-F09",  
    "category": "Personality and behavioural disorders due to brain disease,  
damage and dysfunction",  
    "categoryCode": "F07",  
    "subcategory": "F07.9",  
    "inclusion": [  
        "disorders of psychological development",  
        "Organic psychosyndrome"  
    ],  
    "exclusion": [  
        "symptoms, signs and abnormal clinical and laboratory findings, not  
elsewhere classified"  
    ],  
    "symptoms": [  
        "personality changes",  
        "mood swings",  
        "impaired judgment",  
        "difficulty with impulse control",  
        "altered decision-making ability"  
    ]  
}
```

Данните от СЗО имат и полета с това кои неща се включват и кои не се във въпросния код по МКБ-10, което имаше роля при нагласянето на заявката за търсене (например кое поле с каква тежест да се анализира)

За някои кодове няма подкатегория и затова спира до категория сама, те изглеждат по следния начин:

```
{  
    "icd10code": "A33",  
    "type": "category",  
    "title": "Tetanus neonatorum",  
    "chapter": "Certain infectious and parasitic diseases",  
    "chapterCode": "I",  
    "block": "Other bacterial diseases",  
    "blockCode": "A30-A49",  
    "category": "Tetanus neonatorum",  
    "categoryCode": "A33",  
    "subcategory": "",  
    "inclusion": [  
        "diseases generally recognized as communicable or transmissible"  
    ],  
    "exclusion": [  
        "carrier or suspected carrier of infectious disease",  
        "certain localized infections - see body system-related chapters",  
        "influenza and other acute respiratory infections"  
    ],  
    "symptoms": [  
        "poor feeding",  
        "irritability",  
        "muscle stiffness",  
        "spasms",  
        "difficulty swallowing"  
    ]  
}
```

Като допълнителни данни също бяха добавени:

- EtiologyFactor
- BodySystem

С цел да обогатим повече данните за да можем да правим заявки към тях, колкото се може по-близо до случаи на употреба в реалния живот.

3 Моделиране

Формализъм: OWL 2 DL (Description Logic).

Инструменти: Protégé, WebVOWL.

Моделиране:

- ясно разграничение Клас / Индивид / Свойство
- йерархично моделиране на структурата на МКБ-10 (Chapter → Block → Category → Subcategory)
- дефинирани класове и ограничения (EXISTS / HASVALUE)

`RespiratoryDisease ≡ Disease ⊓ (affectsBodySystem value respiratory_system)`

- дефиниране на взаимно обратни свойства (напр. принадлежност и обратна принадлежност) и транзитивни свойства за навигация в йерархията на МКБ-10;
- поддържане на логическа консистентност и възможност за автоматично извеждане на знания чрез OWL reasoner.

4 Основни класове и йерархия

А. Структурни класове (ICD-10 йерархия):

- Chapter – глава от ICD-10 (напр. I, II, XIX)
- Block – блок в рамките на глава
- Category – категория (код от типа “A00”, “J18”...)
- Subcategory – подкатегория (по-детайлен код)

Category и Subcategory реално ще ги поглеждаме като Disease в рамките на онтологията, понеже има заболявания, които по МКБ-10 са дефинирани само до категория.

Б. Клинични/медицински класове:

- Disease – общ клас за болест/диагностична единица (родителски)
- Symptom – симптом
- EtiologyFactor – етиологичен фактор
- BodySystem – органова система

В. Дефинирани (съставни) класове:

- RespiratoryDisease – дефиниран клас за болести, които засягат дихателната система
- InfectiousDisease – дефиниран клас за болести (глава I)
- NeoplasmDisease – дефиниран клас за неоплазми (глава II)
- InjuryDisease – дефиниран клас за травми (глава XIX)
- SymptomaticDisease – дефиниран клас за болести, които имат поне един симптом

Тези класове са генеририани чрез DL ограничения и след пускане на reasoner индивидите се класифицират автоматично към тях.

5 Обектни свойства (релации)

Свойство	Домейн	Рейндже	Характеристики
<code>affectsBodySystem</code>	<code>Category, Subcategory</code>	<code>BodySystem</code>	<code>inverse: bodySystemAffectedBy</code>
<code>bodySystemAffectedBy</code>	<code>BodySystem</code>	<code>Disease (Category/Subcategory)</code>	<code>inverse: affectsBodySystem (логически, тъй като е обратното)</code>
<code>hasSymptom</code>	<code>Category, Subcategory</code>	<code>Symptom</code>	—
<code>hasEtiology</code>	<code>Category, Subcategory</code>	<code>EtiologyFactor</code>	—

Свойство	Домейн	Рейндже	Характеристики
belongsToChapter	Category, Subcategory	Chapter	—
belongsToBlock	Category, Subcategory	Block	—
hasBlock	Chapter	Block	inverse: blockOfChapter
blockOfChapter	Block	Chapter	inverse: hasBlock (<i>логически, обратното</i>)
hasCategory	Block	Category	inverse: categoryOfBlock
categoryOfBlock	Category	Block	inverse: hasCategory (<i>логически, обратното</i>)
hasSubcategory	Category	Subcategory	inverse: subcategoryOfCategory
subcategoryOfCategory	Subcategory	Category	inverse: hasSubcategory (<i>логически, обратното</i>)
belongsToCategory	Subcategory	Category	subPropertyOf: hasAncestorCategory
hasAncestorCategory	Subcategory	Category	Transitive

6 Свойства по типа на данни

Свойство	Домейн	Тип	Характеристики
codeValue	Chapter, Block, Category, Subcategory	xsd:string	Functional
preferredLabel	Chapter, Block, Category, Subcategory	xsd:string	Functional
title	Chapter, Block, Category, Subcategory	xsd:string	—
description	Category, Subcategory	xsd:string	—
rangeStart	Block	xsd:string	—
rangeEnd	Block	xsd:string	—

7 Индивиди и извеждане (Reasoning)

Примери за автоматични изводи

- **Респираторни заболявания:** от affectsBodySystem(J18, respiratory_system) \Rightarrow J18 \in RespiratoryDisease.
- **Инфекциозни заболявания:** от belongsToChapter(Aoo, I) \Rightarrow Aoo \in InfectiousDisease.
- **Симптоматични заболявания:** от hasSymptom(J18, cough) \Rightarrow J18 \in SymptomaticDisease.
- **Йерархия на МКБ-10:** от belongsToCategory(J18.1, J18) и belongsToCategory(J18, J) \Rightarrow hasAncestorCategory(J18.1, J).

8 Компетентностни въпроси и заявки

Компетентностни въпроси

1. Кои заболявания имат даден симптом (напр. кашлица)?
2. Кои заболявания се характеризират с комбинация от симптоми (напр. температура и кашлица)?

3. Кои заболявания засягат определена органова система (напр. дихателната система)?
4. Кои заболявания са свързани с конкретен етиологичен фактор (напр. вирусен причинител)?
5. Кои заболявания засягат дадена органова система и едновременно с това проявяват конкретен симптом?

DL заявки (Manchester syntax)

Disease and (hasSymptom value cough)

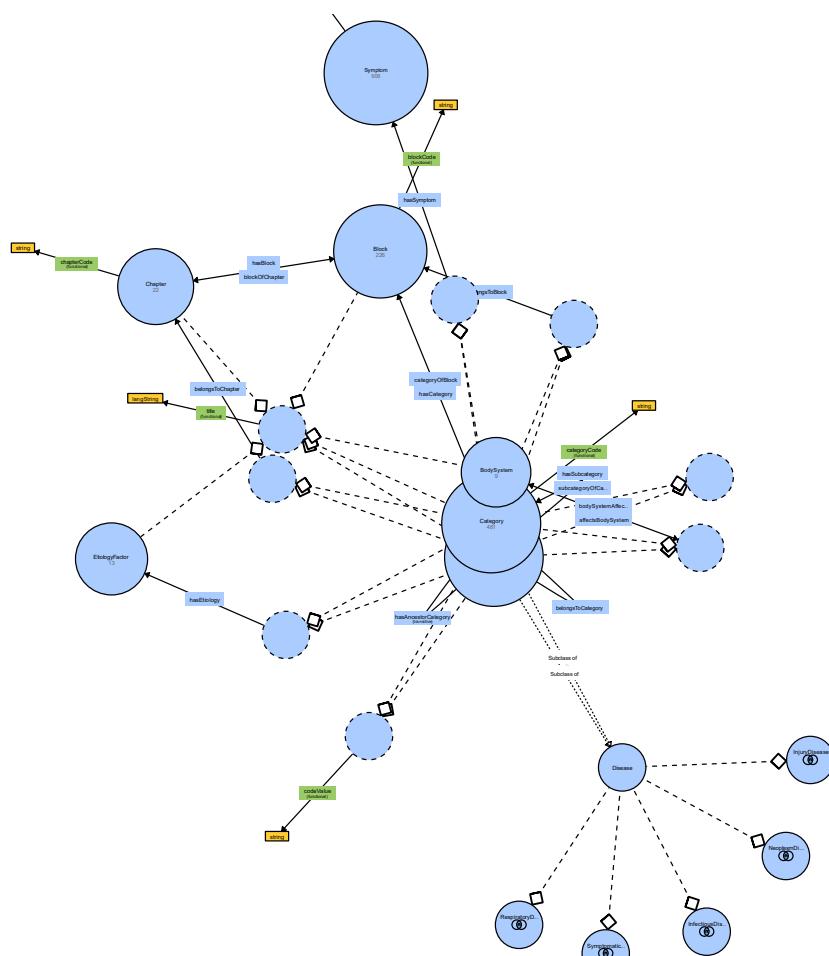
Disease and (hasSymptom value fever) and (hasSymptom value cough)

Disease and (affectsBodySystem value respiratory_system)

Disease and (hasEtiologyFactor value viral_agent)

Disease
and (affectsBodySystem value respiratory_system)
and (hasSymptom value fever)

9 Визуализация (WebOWL)



Фигура 1: WebOWL Визуализация

10 Бъдещо развитие

Следващи стъпки:

- допълнителни клинични характеристики, като стадий на заболяването, хроничност и тежест
- интеграция на по-детайлни етиологични и рискови фактори
- добавяне на времеви аспекти (начало, продължителност, рецидиви) при описанието на заболяванията
- интеграция на онтологията в медицински информационни и експертни системи

11 Съдържание, генерирано с изкуствен интелект

- части от кода, който генерира онтологията от данните в .owl файл
- правилен синтаксис за DL заявки

12 Използвана литература

- [1] МКБ, 10-та ревизия (МКБ-10), [<https://icd.who.int/browse10/2019/en>]
- [2] кодове на МКБ-10, клинична модификация, [<https://github.com/k4m1113/ICD-10-CSV/blob/master/codes.csv>]
- [3] ICD API, [<https://icd.who.int/icdapi>]
- [4] Набор от данни 1, [<https://www.kaggle.com/datasets/dhivyeshrk/diseases-and-symptoms-dataset/data>]
- [5] Набор от данни 2, [<https://www.kaggle.com/datasets/phoenion/diseases-and-their-symptoms>]
- [6] Набор от данни 3, [<https://www.kaggle.com/datasets/choongqianzheng/disease-and-symptoms-dataset>]
- [7] Набор от данни по Disease Ontology, [<https://github.com/dhimmel/hsdn/blob/gh-pages/data/symptoms-DO.tsv>]
- [8] Disease Ontology mapping,
<https://github.com/DiseaseOntology/HumanDiseaseOntology/blob/main/DOreports/allXREFinDO.tsv>