Предложения за теми за курсови проекти по Изкуствен интелект

Около 80% от медицинската документация се съхранява в текстов формат. За обработката и извличане на знания в структуриран вид е необходимо създаването на специални алгоритми от ИИ за обработка на медицински текст.

Ще се използват корпуси от

1. Откриване в клинични записи на пациенти на описания на минали заболявания и тяхната давност. Зададена е база с анамнези на пациенти:
   1. Идентифициране в анамнезата (история на заболяването) на имена на диагнози и генериране на списък със съответните им кодове по международната класификация на болестите (ICD-10). Може да се използва като база корпус от 350000 текстови описания на диагнози с анотирани кодове: <https://github.com/BorisVelichkov/ICD10-Medical-Data>
   2. Идентифициране в анамнезата (история на заболяването) на оплаквания/симптоми. Може да се използва като база корпус генериран от Wikidata (<https://www.wikidata.org/wiki/Q16987> , <https://w.wiki/i9n> ), както и от данните в <https://www.puls.bg/diagnostic/symptom/>
   3. Идентифициране в анамнезата (история на заболяването) на имена на диагнози и генериране на списък със съответните им кодове по международната класификация на болестите (ICD-10). Може да се използва като база корпус от 350000 текстови описания на диагнози с анотирани кодове: <https://github.com/BorisVelichkov/ICD10-Medical-Data>
   4. Идентифициране в анамнезата (история на заболяването) на описание на предходно лечение. За откриване на имената на лекарства може да е използва лекарствен списък на НЗОК (<https://www.nhif.bg/page/45> )
   5. Идентифициране в анамнезата (история на заболяването) на рискови фактори, като:
      1. пушене
      2. злоупотреба с големи количества алкохол
      3. приемане на вредни субстанции и др.
   6. Идентифициране в текст на описания на периоди от време (темпорални единици): дати, времеви период и др. – експлицитно или относително описани и тяхно описание според TIMEX3 (<http://www.timeml.org/publications/timeMLdocs/timeml_1.2.1.html> )
2. Откриване в клинични записи на пациенти на тяхното текущо състояние. Зададена е база с описания на статуси на пациенти:
   1. Идентифициране в статус на пациент на стойности за неговото артериално кръвно налягане и пулс
   2. Идентифициране в статус на пациент на стойности за неговите ръст, тегло, ИТМ (BMI – body mass index), обиколка на талия
   3. Идентифициране в статус на пациент на стойности за неговата телесна температура
   4. Идентифициране в статус на пациент на анатомични органи и системи, които са без увреждания, т.е. „б.о.“
   5. Представяне на данни в структуриран вид за различни органи и системи:
      1. Сърдечно съдова система
      2. Корем
      3. Глава
      4. Кожа
      5. Дихателна система
      6. Черен дроб
      7. Слезка
      8. Опорно-двигателен апарат
      9. Пикочно-полова система
      10. И др.
3. Откриване на тенденции, сложни взаимовръзки между заболявания, региони, възраст и пол на пациенти и време. Зададена е база със следните данни:

[PatientID; Date; RegionID, Age, Gender, ICD10, ICD10, ICD10, ICD10]

Където броят на диагнозите по ICD10 може да варира между 1 и 4

* 1. Откриване на коморбидности – сложни взаимовръзки между едновременна поява на заболявания, които на пръв поглед не са свързани
  2. Тенденции за разпространение на заболявания
     1. По региони
     2. Във времето
     3. Като следствие от минали заболявания
     4. По възрастови групи
  3. Откриване на хронология на заболяванията на пациентите и често повтарящи се групи от диагнози във времето
     1. Като предварително се маркират хроничните заболявания и се изключат от веригата.
     2. Изследване на предходни заболявания преди появата на диагноза на хронично заболяване, в т.ч. и други хронични заболявания