

The medTurk Book



July 11, 2014

Preface

ICBI at Georgetown University is proud to provide this simple, yet effective, tool for converting unstructured clinical notes into structured data for clinical research. This book contains instructions on how to set up and use medTurk.

- Robert M. Johnson (rmj49@georgetown.edu)

Contents

1	Introduction	1
2	Installation	3
2.1	Python	3
2.2	UMLS	3
2.3	cTAKES	3
2.4	mongoDB	4
3	Creating a Toy Example Project	5
3.1	Loading Clinical Data	5
3.2	Loading	5
3.3	Installing cTAKES	5
3.4	Installing UMLS	5

Chapter 1

Introduction

medTurk is web application that aims to coordinate the activity of mutiple users to convert unstructured clinical notes into structured clinical data for research. medTurk assumes there are questions you would like answered about your patients' data. To encapsulate these questions, types of allowable answers, medTurk uses 'Research Models' or RMs for short. A RM is a simple JSON file that organizes questions, answers, and taxonmies.

It works as follows. First, you must convert your clinical notes into a simple JSON format as described in Chapter 2. Next, you process these clinical notes using cTAKES. After this, you create a project

Chapter 2

Installation

Since we have yet to provide a simple installer package for medTurk, in this section we guide you through the process of setting up medTurk. medTurk depends on Python and several Python packages. In addition, it requires cTAKES for processing clinical notes, a local running instance of UMLS, and a local running instance of MongoDB. This chapter instructs you on how to setup each of these components. We start with Python dependencies.

2.1 Python

medTurk requires a Python version of at least 2.7.6. On a Mac Terminal, you can use following line to retrieve the current Python version:

```
python --version
```

Next, using pip, you must install the following Python packages:

```
pip install Flask
pip install mimerender
```

2.2 UMLS

medTurk uses UMLS to look up names associated with CUIs reported by cTAKES. The following webpage provides instructions on how to setup a UMLS MySQL database locally:

http://groups.csail.mit.edu/medg/projects/text/Load_UMLS_mysql.html

2.3 cTAKES

medTurk uses cTAKES to process each clinical note for concepts and stores ones which meet three following criteria. The first criteria is negation. The concept must have positive polarity, that is, it is not negated (e.g. the patient denies smoking).

The second criteria is the concept must have a subject of patient. As an example, there are times when the a physician might write that a patient's father had lung cancer. In this example, the concept of "lung cancer" would have a subject of "Patient's Father". medTurk only extracts concepts which act directly on the patient.

The last criteria is confidence. cTAKES reports a confidence number for each extraction and medTurk only keeps concepts which have a confidence value of "1.0".

To install cTAKES, their website provides instructions on how to do this:

<https://cwiki.apache.org/confluence/display/CTAKES/cTAKES+3.1+User+Install+Guide>

2.4 MongoDB

To install MongoDB, following the instructions on this webpage:

<http://www.mongodb.org/downloads>

Chapter 3

Creating a Toy Example Project

This chapter gets you started by running a toy example that is provided by medTurk. This toy example provides clinical notes and a research model to get you started. For your real project, you will of course load your own data and research model, but the steps remain the same nevertheless.

3.1 Loading Clinical Data

3.2 Loading

dsfsdf

3.3 Installing cTAKES

dsfsdf

3.4 Installing UMLS

dsfsdf