CS 682-01: Software Development Lab Projects
Fall 2020

1 BostonApartments.com

Admin: Eric Boyer, Boston Apartments.com, webmaster@bostonapartments.com, +1 617-254-5501

Description: BostonApartments.com features real estate agencies and landlords with apartments in Boston, throughout Massachusetts, Southern New Hampshire and in Rhode Island. It is a full featured site including real estate for sale, commercial real estate, a roommate service, advice and advertising from real estate related services and of course apartments for rent. People come to Boston Apartments.com looking for apartments in Boston and various other topics and help that is available on this oldest and longest running real estate site for apartments in Boston. The classified ads on Boston Apartments.com are constantly updated by the participating real estate agencies that use BostonApartments.com's proprietary apartment database to run their real estate offices' apartment listings. This specialized software allows these companies to keep their listings for apartments in Boston and beyond up-to-date. BostonApartments.com is more than just apartments in Boston.

We're looking to have students evaluate existing applications to recommend improvements and design and develop a new application. There are 2 types users of BostonApartments.com. Those that read it and those that use it as a back end system to run their websites, syndicate to other sites, posts to social media, etc. The current server is an older Debian version, upgrades to interface, server, software need to be done. Some functionality may be affected due to the age of the website upon upgrading the server as some of the commands may be deprecated or just inefficient with the changes in language. Mobile applications and modernization/of the CrM portion of the website (agent to clients). Updating the social network posting routines and the chrome extension to post to Craigslist.

Skills and Technologies Required:

- Web design.
- Web application development.
- PHP 7.
- Database design.

Resources: These resources will be provided by the client:

- Access to server, dedicated URL.
- Other resources can be made available upon request.

2 Lendica Data Connectivity

Admin: Jerry Shu, shubuliao@gmail.com

Description: Lendica is a Boston-based financial technology startup that creates smart underwriting technology to provide small and mid-size businesses with easy access to working capital. Lendica enables large institutions and boutique originators to instantly qualify and issue debt capital. Lendica and its partners service all types of businesses including high growth/emerging industries and those recovering from COVID impact.

For this project, students will work on connecting Lendica's underwriting engine with different technology partners through API/Web scraping/etc.

Skills and Technologies Required:

- Python/web scripting
- Good understanding of OAuth 2.0

Resources: Resources can be made available upon request.

3 Named Entity Recognition and Corpus Extraction*

Admin: Mitch, JeevaHealth.ai, mitch@jeevahealth.ai

Description: This project has some machine learning component and it is in two folds:

- Named Entity Recognition (NER) and pronoun resolution:- The goal of this project is to implement a named entity recognizer (for people and events, places) plus a (probably separate) pronoun dereferencer that infers the identity of names that are represented by pronouns. The goal is for it to extract named things in a conversation (like events and people) that are relevant to the previously mentioned corpora for mental health dialogs. This is not traditional off-the-shelf NER that categorizes items but instead extracts the individual items. Events may in the end be a summary of a sentence.

  Final deliverable is an NLP pipeline in Python that when run on sample sentences accurately extracts named entities and identifies pronoun references, plus a record of modifications of the hyperparameters/architecture/process/model and associated accuracy stats for each stage in the sequence of changes.
- Corpus creation and tagging:- The goal of the project is corpus curation: exploration, selection, and tagging lots and lots of text that can be used as a corpus for training the ML algorithm (either for a full language model or refining an existing model with transfer learning). We're looking for text of multiturn conversations for dealing with emotions and mental health in a therapeutic manner, eg conversations discussing emotional problems and how managed (actual therapy transcripts would be best but that is very difficult to find). Avoids tendentious, toxic, shaming, or trolling language. For example, some subreddits are toxic but some subreddits involve discussions that are not at all toxic. There should be multiple registers of language: some more informal and real world, uses abbreviations, colloquialisms, txtspeech, emojis, etc and some more formal like more therapy or helpful responses.

Skills and Technologies Required:

- Python
- Natural Language Processing (NLP)
- AI

Resources: Resources can be made available upon request.

* Students will be required to sign a non-disclosure agreement.