# Dictation Evaluation Reddit Parser

Garcia, Benjamin  Lembke, Logan  MacMillan, Kyle
Smith, Christopher  Stelter, Andrew

September 21, 2018

# Contents

# 1 Introduction to DERP

Dictation Evaluation Reddit Parser – DERP, is designed to be a Domain Specific Language – DSL for interpretting news, or news-like websites. The sysyem was designed as part of a group project for South Dakota School of Mines & Technology's Compilers course.

## 1.1 What is DERP: Human Queries For a Digital World

Time is something we are regularly lacking and it would be nice to expedite the intake of online news while doing other things such as prepparing and eating breakfast, taking a shower, getting prepared, or driving to work. We should be able to tell our phone, "What's going on in the world today?" and be able to parse through available news stories. If we want news on a specific topic such as Tesla we should be able to say, "What's in the news about Tesla?".

In the current form of news intake we either get it through a radio or television broadcast, an article, or a newspaper. The broadcast is rigid and we have no control over what news we hear and there are frequently topics that don't concern us. Articles and newspapers are the same way but we are able to skip what we don't like. The issue with newspapers is we are limited to what the editor decided to include. The issue with articles is tracking down relevant articles. A common way to track down interesting articles is an aggregation site such as reddit, steemit, band, and voat to name a few.

## 1.2 DERP Origins

In the fall of 2018 we were tasked with making a DSL for a class in compiler theory. We discussed many possible topics:

- Constrain/Generate
  - Floor Plans
  - Workout Assistant
  - Computer Part Picker
- Perform a Task
  - Robot "AI" Proposal by Dr. Hinker
  - Simple Image Processing
- Configure a Task
  - News Reader
  - SQL Helper

We decided against a Constrain/Generate model because the ones we came up with all required labeled data which we did not have access to and were not sure we could find or make in a timely manner.

Performing a task seemed like a good idea but the idea of being reliant on getting a robot (in the case of Dr. Hinker's proposal) seemed like a bad idea. We were also unsure of the code that went into it and if we'd be walking into a well documented and fleshed out product or a hack. The image processing was to be something along the lines of simplistic image manipulation.

Lastly we discussed configuration-style tasks. The two proposals were News Reader and an SQL Helper. The basic idea with the news reader is that a user could create custom queries that would obtain news that he/she was interested in while skipping the rest. SQL Helper was an idea to have anyone use SQL. So if a businessman needed some form of report he could ask his secretary to make it and she'd be able to without knowing SQL. The headache of building a system that was robust enough to translate human queries into SQL was deemed too much for the short amount of time we had and so we settled on the News Reader.

We needed a name. Having selected our project we could now determine what to name our team. We settled on Dictation Evaluation Reddit Parser because if we have enough time we'd like it to take spoken words to evaluate. The reddit bit was added because our minimum viable product will be built around reddit.

## 1.3 Why DERP

TODO

# 2 DERP Design Goals

## 2.1 Consolidated Information

The DERP project provides an intuitive way to multiplex multiple news sources into a personalized stream of data. Using DERP, one can specify as many different sources as they want to (provided language plugins are available), and then all of those sources are accessed through the same set of language keywords, regardless of if the source is a subreddit, a news website, or just a file on the user's device. Furthermore, DERP facilitates naming groups of sources and queries to create query macros, allowing users to further personalize what they get out of DERP.

Online news sources regularly expose the same categories of information - date, author, title - DERP acts as an interface for all different article types, allowing users to work only with this high-level information about the articles they are finding. Hiding the details of what exactly is required to find an article that matches a specific criterion allows users to focus more on what they want to read, and less on how they obtain that information.

## 2.2 Easy to Use Natural English Interface

The second goal of DERP is to design a language that won't feel strange to speak. All of the DERP keywords and syntax are similar to natural English speech patterns. This allows for easy adaptation of DERP into speech-recognition tools such as Google Assistant and Amazon Alexa. Using the natural form of the language, users with one of these devices could speak their program to the interpreter and receive their results immediately.

In addition to being easy to command DERP through natural language, DERP provides output that also feels natural. DERP has a set of phrases available to it for reporting errors or results from user queries. The main output from DERP, articles the user has requested, are outputted as full text so that devices reading them using a text-to-speech system sound as natural as possible.

## 2.3 Search More, Search Faster

By amalgamating news sources into a personalized feed, DERP allows users to find the information they want with fewer operations. Rather than check each of their favorite subreddits, news sites, and RSS feeds, a user can simply load those sources into DERP and make a general query about them. DERP will do the heavy lifting of finding things the user might be interested, which means the user will be able to spend more of their time actually consuming the information provided and deciding on new topics to get information over.

Because it is an extensible language, DERP users are limited only by the language plugins they use. While some keywords are understood specially by the language, such as those pertaining to dates and times, any language plugin

can provide additional fields and keywords, making the language flexible and powerful while keeping the complexity out of the core of the language.

# 3 The DERP Execution Model

**3.1 Interactive**

**3.2 Hierarchical Data (website → boards (subreddits) → posts**

**3.3 Constraints**

**3.4 Savable Criteria (data structures)**

**3.5 Composed Constraints**

# 4 The DERP Query Language

## 4.1 Main Features

### 4.1.1 Primitives

### 4.1.2 Higher Order Types...

### 4.1.3 Memory Management

## 4.2 DERP Operators