

COMP SCI 349 - Final Project
User-Generated RPG Content: An NLP Task (Option #4)

Part 1: The Problem

For this project I chose to pursue Option #4 using a data set I curated myself using the Python Reddit API Wrapper (PRAW). Briefly, I collected several thousand links to fan-generated role playing game content designed for Dungeons and Dragons 5th edition and its Open Gaming License derivatives from the */r/UnearthedAcana* and */r/DnDHomebrew* subreddits, including post information such as user voting scores and post flairs (user-specified tag keywords).

For simplicity's sake, I targeted posts with links to the following domains: *homebrewery.naturalcrit.com* and *gmbinder.com* as these websites provide a useful markdown-style notation for creating this type of content that makes collecting the relevant text more straightforward. Examples of raw and rendered content taken from the recently-posted public content of *gmbinder.com* can be found below.

Bodysmithers Specialist

There are works that draw fury from the nobility, those that revolt temples, and others that drive academics crazy. The art of bodysmithing outrages all of them. It is not an ungrateful labor, however. Not by a long shot. The bodysmith provides for those in need of accessibility and the mad overachievers alike. The bodysmith works the same, but no matter if their craft takes the form of prosthetic machines, bio-enhancements or magic-imbued primal elements, the bodysmith takes fulfillment on the perfection of bodies and the amazing feats anyone with enough drive (and bodily humours) can accomplish.

Tools of Physiology

3rd-level Bodysmither feature

You gain proficiency with the Medicine skill. If you already have this proficiency, you gain proficiency with one other skill or tool of your choice. You know how to create, repair and apply prosthetics using the Medicine skill or tool of relevance.

Bodysmithers Spells

3rd-level Bodysmither feature

You always have certain spells prepared after you reach particular levels in this class, as shown in the Bodysmithers Spells table. These spells count as artificer spells for you, but they don't count against the number of artificer spells you prepare.

Artificer Level	Spell
3rd	<i>fog cloud, magic missile</i>
5th	<i>barkskin, shatter</i>
9th	<i>animate dead, call lightning</i>
13th	<i>greater invisibility, locate creature</i>
17th	<i>modify memory, Rary's telepathic bond</i>

Maintenance Hobbyist

3rd-level Bodysmither feature

You know how to create, apply and maintain prostheses, and it's something you delight off during your free time! You may work on a number of prosthetics and Augmentations equal to double your proficiency modifier while taking short or long rests. For the purposes of resting, maintenance work is considered a light activity for you, and you don't need to make skill checks to maintain a prosthetic. You can use a maintained prosthetic on yourself as a spellcasting focus for your artificer spells, and can also apply your Magic Tinkering features to them.

You can create Disposable or Permanent prosthetics that simulate natural movement. A Disposable prosthetic can be created during a Short Rest expending 50gp of materials at hand. A disposable prosthetic breaks completely at the beginning of the user's next long rest. You may then recreate it, expending more 25gp in materials. To build a Permanent prosthetic you must expend 500gp in materials during a Long Rest. You can create only one prosthetic per rest.

ADAPTATION

Despite the bodysmith work being (mostly) carefully done to avoid complications, the process of adaptation to new prosthetic limbs is still tough. The user takes 1 point of exhaustion per new prosthetic at application, and at the end of each day, for 1d6 days. The user may also experience disadvantage on activities that require effort or finesse for 1d4 days from implementation.

Additionally, whenever a character receives a prosthetic addition, roll 1d6, or choose, between the following recurring side-effects:

d6	Loot
1	The user loses sensation of the prosthetic area
2	Pain is heightened on the prosthetic area
3	Weather may cause the prosthetic to creak or jam
4	The user might need to concentrate to control the prosthetic's strength
5	The area of the prosthetic may get infected from time to time
6	The prosthetic may sometimes move by its own

Figure 1: Rendered web-view <https://www.gmbinder.com/share/-MbD4VSGfFzhW7aaNfHw>

Bodysmithers Specialist

There are works that draw fury from the nobility, those that revolt temples, and others that drive academics crazy. The art of bodysmithing outrages all of them. It is not an ungrateful labor, however. Not by a long shot. The bodysmith provides for those in need of accessibility and the mad overachievers alike. No bodysmith works the same, but no matter if their craft takes the form of prosthetic machines, bio-enhancements or magic-imbued primal elements, the bodysmith takes fulfillment on the perfection of bodies and the amazing feats anyone with enough drive (and bodily humours) can accomplish.

Tools of Physiology

3rd-level Bodysmither feature_

 You gain proficiency with the Medicine skill. If you already have this proficiency, you gain proficiency with one other skill or tool of your choice. You know how to create, repair and apply prosthetics using the Medicine skill or tool of relevance.

Bodysmithers Spells

3rd-level Bodysmither feature_

 You always have certain spells prepared after you reach particular levels in this class, as shown in the Bodysmithers Spells table. These spells count as artificer spells for you, but they don't count against the number of artificer spells you prepare.

Artificer Level	Spell
3rd	<i>fog cloud, magic missile_</i>
5th	<i>barkskin, shatter_</i>
9th	<i>animate dead, call lightning_</i>
13th	<i>greater invisibility, locate creature_</i>
17th	<i>modify memory, Rary's telepathic bond_</i>

Maintenance Hobbyist

3rd-level Bodysmither feature_

 You know how to create, apply and maintain prostheses, and it's something you delight off during your free time! You may work on a number of prosthetics and Augmentations equal to double your proficiency modifier while taking short or long rests. For the purposes of resting, maintenance work is considered a light activity for you, and you don't need to make skill checks to maintain a prosthetic. You can use a maintained prosthetic on yourself as a spellcasting focus for your artificer spells, and can also apply your Magic Tinkering features to them.

You can create Disposable or Permanent prosthetics that simulate natural movement. A Disposable prosthetic can be created during a Short Rest expending 50gp of materials at hand. A disposable prosthetic breaks completely at the beginning of the user's next long rest. You may then recreate it, expending more 25gp in materials. To build a Permanent prosthetic you must expend 500gp in materials during a Long Rest. You can create only one prosthetic per rest.

> ##### Adaptation

> Despite the bodysmith work being (mostly) carefully done to avoid complications, the process of adaptation to new prosthetic limbs is still tough. The user takes 1 point of exhaustion per new prosthetic at application, and at the end of each day, for 1d6 days. The user may also experience disadvantage on activities that require effort or finesse for 1d4 days from implementation.

Additionally, whenever a character receives a prosthetic addition, roll 1d6, or choose, between the following recurring side-effects:

d6	Loot
1	The user loses sensation of the prosthetic area
2	Pain is heightened on the prosthetic area
3	Weather may cause the prosthetic to creak or jam
4	The user might need to concentrate to control the prosthetic's strength
5	The area of the prosthetic may get infected from time to time
6	The prosthetic may sometimes move by its own

<pagebreak

Figure 2: Markdown source https://www.gmbinder.com/share/-MbD4VSGfFzhW7aaNfHw/-MbDwIj2eYXtUtYk-YF_/source

Neil Kuehnle

NetID: njk639

Repository: https://github.com/njk639/COMP_SCI_349_MachineLearning_Final

Part 2: Defining Objectives

Outside of my core academic research, I've dabbled in building an application that allows users to navigate content and create "player characters" for several different RPG titles. To this end, I've tried to create a framework for representing the mechanical details of RPG content in a flexible, easily-customized way. A useful machine learning application I'd like to attempt integrating into this is a tool which can determine the type of content contained in third-party text content shared by users without requiring the users to manually generate the YAML configuration files I use to represent resources.

Long-term this is a multi-class/multi-label segmentation problem, but initially I have two goals in mind:

1. Create a simple classifier that can categorize content based on broad classes epitomized by Reddit link flair.
2. Perform unsupervised clustering of content as a starting point to help simplify the manual labeling of mechanically-relevant spatial labels.

Part 3: Exploratory Data Analysis and Narrowing-Down Problem

I decided to look at flair class balance on */r/UncarthedArcana*, where the flair categories are strictly linked to the type of content linked.

This revealed a highly-skewed set of classes, with flairs like background, compendium, feat, feature, mechanic, official, other, prestige, resource, and world being extremely uncommon. Based on this, I decided to exclude all but the following classes when training a post flair classifier. Class, Item, Monster, Race, Spell, and Subclass.

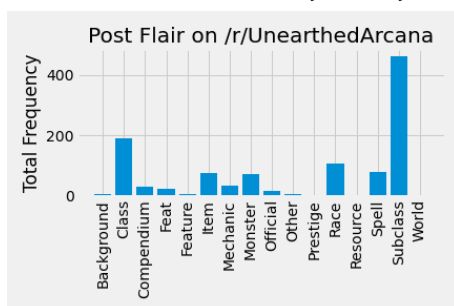


Figure 4: Total frequency of classes in */r/UncarthedArcana* dataset prior to filtering.

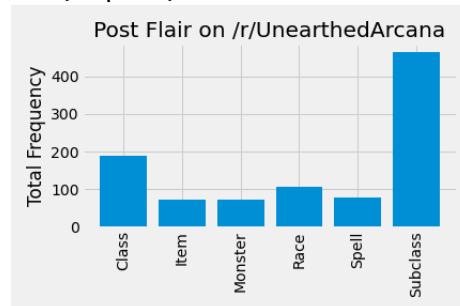


Figure 3: Total frequency of classes in */r/UncarthedArcana* dataset after filtering for the 6 most-common classes.

Note: that the classes are still imbalanced, which I will be addressing later on.

Part 4: Cleaning Data and Exploring Classes

I initially filtered and cleaned my data set as much as possible using a large number of regular expressions to remove various erroneous characters (lone instances of 'Â' and 'Z', etc.), CSS carried over despite BeautifulSoup4's text rendering function, and elements of the markdown, such as | and _ characters.

Following this, I pooled all of the text from all instances of each class and plotted their term frequencies using the WordCloud package.

By far the most common term associated with the 'Class' flair was level which is likely related to the frequent use of the word level to indicate when different elements of a character's class are awarded to them.

This was also similar for the 'Subclass' flair, which as the name implies is a specialized subset of the 'Class' flair. However, terms associated with some of the specific features granted by a subclass appear more prominently than in the main 'Class' category, possibly because these tend to be more strictly focused on mechanical features granted to a

Neil Kuehnle
NetID: njk639

Repository: https://github.com/njk639/COMP_SCI_349_MachineLearning_Final

character player. Thus terms like action, spell, feature, attack, and turn are all more common.

For the 'Spell' flair tags, words such as spell, action, creature, V/S (abbreviations for verbal/somatic components which describe elements required to use spells under the DnD 5E rules) are all common phrases used to describe either in-game abilities/actions, including spells.

For the 'Monster' category we see several term associated with the statistics and various in-game abilities possessed by 'Monsters' (attack, action, hit, target, damage, throw, etc.) as well as the common in-game 'Monster' synonym: creature.

Similarly, the 'Item' category is enriched for terms associated with the mechanical features of different potential items, such as weapon, damage, action, creature, charge, hit, attack, rare, target, etc. While the most common term is 'item.'

Lastly, the 'Race' (fantasy species: elves, dwarves, etc.) category, has several terms associated with different mechanical features as above (action, spell, etc) but also emphasize the terms 'level' and 'trait' which is likely reflective of the common DnD 5E syntax of racial mechanical features as 'traits' which are often associated with a level-based award system as in classes/subclasses. Of note is the high frequency of the term "dragon." This raises some initial concerns for me about potential bias within that dataset, given its relatively small sample size (86 examples) it is possible that the concept of "dragon"-based "races" are over-sampled within my data, which may lead to it under-performing on models that do not include the "dragon" concept within them.

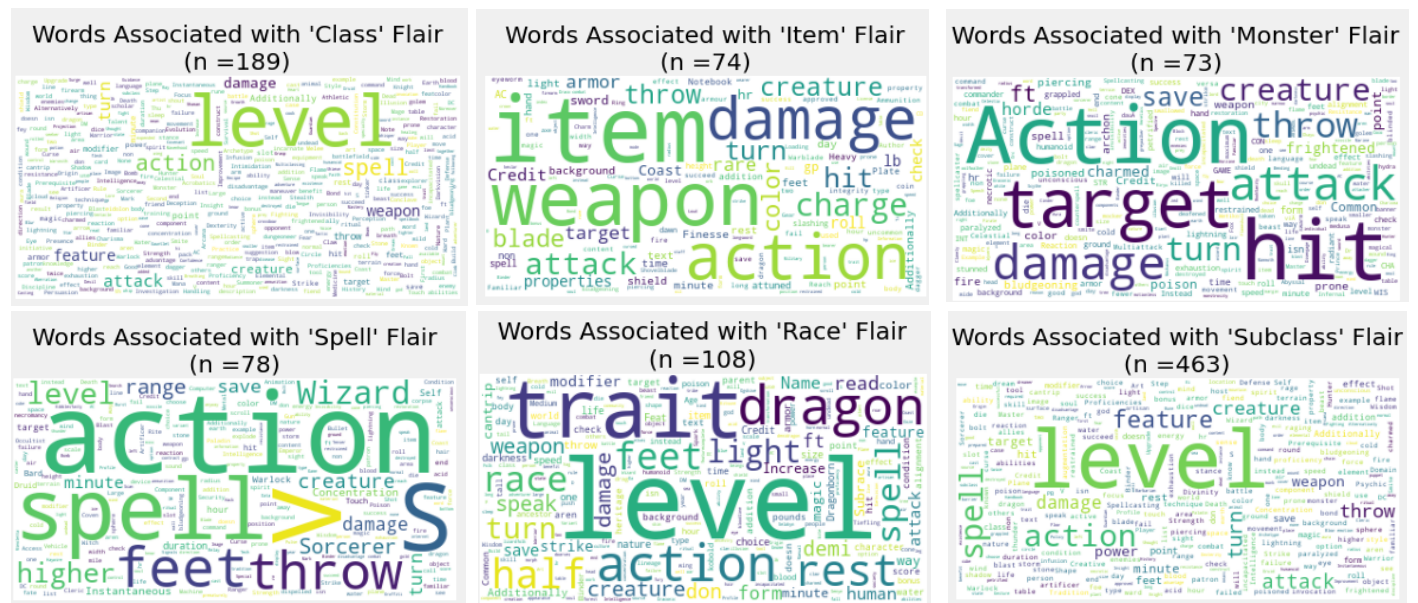


Figure 5: Word clouds of single (1-gram) words present in the aggregate corpus of each filtered class set. Size is directly proportional to frequency within the class overall.

Part 5: Building a classifier

I felt that my training sample was already quite small relative to the potential dimensions of my problem. Since my data was originally collected 3 months prior before I began this project, I decided to collect the 1000 most recent posts from /r/UnearthedArcana to build a testing data set.

Neil Kuehnle

NetID: njk639

Repository: https://github.com/njk639/COMP_SCI_349_MachineLearning_Final

After this I ended up with a training set of 795 examples and a testing set of 314 examples. The exact class label breakdowns can be found below, but appears largely similar between the two dataset splits.

Split	Flair	Count
Test	Class	52
	Item	16
	Monster	18
	Race	22
	Spell	17
	Subclass	163
Train	Class	137
	Item	58
	Monster	55
	Race	86
	Spell	61
	Subclass	300

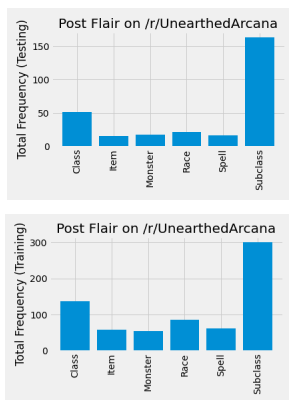


Figure 6: Class composition for testing and training splits.

To build this model, I imported Sci-Kit learn. I first made use of the CountVectorizer to transform my corpus into a document-term matrix—I chose not to use the TF-IDF (term frequency-inverse document frequency) vectorizer. Normally this might be advantageous since it normalizes the term frequencies to the inverse log of their overall frequency within the entire corpus of documents; however, I felt it was ill-advisable due to my severe class imbalance. It's also worth noting that something like NLTK would be better suited for tokenizing/vectorizing my word counts.

I further filtered this to consider only words that had greater than 5 total instances throughout my entire corpus. This reduced the overall dimensions of my problem from 20,807 terms to 5,700. This should help limit the number of highly-document specific terms (i.e. proper nouns such as invented names or author bylines) and in theory should help the model generalize slightly better without being too restrictive on the classes with approximately 50 examples (as only 10% would need to contain the term once to qualify). Nonetheless, this type of manual threshold would benefit from additional experimental tuning. In the end, I decided to further reduce my dimensionality by applying PCA until I captured 99% of the cumulative variance and trained my model using the 321 principle components returned (again this might be an area for future tuning).

Next, I made use of the one-vs-rest multi-classification framework in sklearn in conjunction with a simple SVC model, setting the class weights set to 'balanced' (each class is weighted by its relative abundance among all examples) as I was still worried about my model being unduly biased towards certain over-enriched classes (e.g. the 'Subclass' flair).

Overall, my model achieved between 80-100% training accuracy on a per-class basis, with 80-95% testing accuracy for all but one class ('Spell' flair), where accuracy dropped from 86.89% on the training data to 58.82% on the testing data. I think this is promising for a simple classifier without any spatial information whatsoever available. It is likely that including n-grams of >1 would further improve accuracy by mapping features to paired sets of terms. However, should I pursue this project further, I would aim to develop a convolutional neural network or even a transformer architecture. This would also provide a necessary architecture to exploring options for applying multi-label classification and segmentation to annotate larger documents, such as books or compendiums of in-game resources.

Neil Kuehnle

NetID: njk639

Repository: https://github.com/njk639/COMP_SCI_349_MachineLearning_Final

My overall (macro-averaged) testing metrics are as follows: accuracy = 0.861, precision = 0.805, recall = 0.824, F1 score = 0.799.

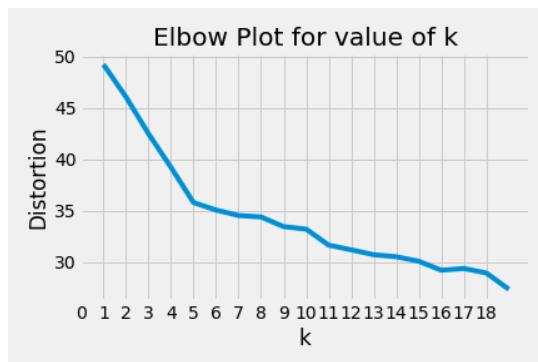
Individual class metrics are found below:

	Training Accuracy	Testing Accuracy	Precision	Recall	F1 Score
Class	0.839	0.826	0.811	0.826	0.819
Item	0.982	0.937	0.468	0.937	0.625
Monster	0.854	0.833	0.789	0.833	0.81
Race	0.720	0.863	1	0.863	0.926
Spell	0.868	0.588	0.833	0.588	0.689
Subclass	0.873	0.895	0.945	0.895	0.924

Part 6: Unsupervised Clustering of Text

Lastly, I briefly wanted to test some clustering of different types of text. Ultimately, I'd like to do this on individual paragraphs/sections of each document as denoted by line breaks and/or '#' characters in the markdown as a way to help guide some more sophisticated and meaningful spatial labeling of the source text.

I started by performing k-means clustering using the elbow method to compare distortion vs the number of clusters across content taken from both */r/UncarthedArcana* and */r/DnDHomebrew*.



There are 16 types of flair utilized on */r/UncarthedArcana* (where the required flair labels provide human-meaningful context). However, only 11 of these would be flair that I would consider broadly applicable to the type of content that has true mechanically relevant meaning to most players (Features/Feats combined to a single category and excluding official, prestige, resource, world flair). Interestingly, 11 seems like a fairly decent (perhaps slightly overfit) level of clustering for my data.

I decided to generate a Guassian mixture model with 11 clusters using my entire dataset and then see what labels these would have when labeled by their most frequent flair from among the */r/UncarthedArcana* subset. Overall this naive approach seems to have not performed particularly well, as all flair were evenly distributed (Chi-Square test). However, it may still present value if performed at the paragraph level in order to cluster more basic/primitive mechanical elements present at the paragraph level.