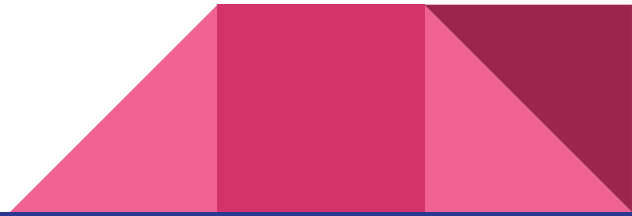
WikiClassify

Nathan Kjer, Luke Wielgus, Adam Massoud, Brian Faure, Wanshu (Wayne) Sun, Viswanathan Subramanian, Brian Chu

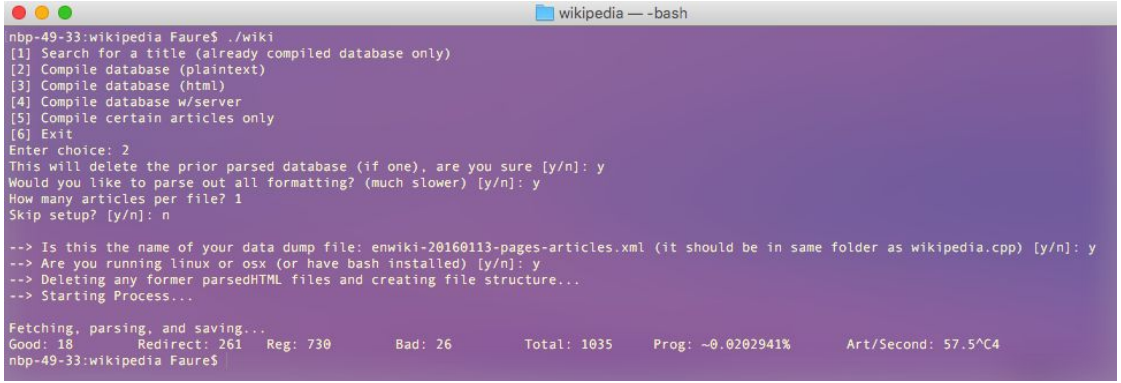
Website/Chrome Extension

- User Interface will be done through a simple website and chrome extension
  - Search bar
    - User can search for desired Wikipedia article
  - Link to About and Login
- Hardware requirement: internet connected device



Parsing Database

- All Wikipedia data is stored in one central dump
- We parse through this dump splitting up articles
- Database stores all the articles and allows search functionality



Machine Learning

- System is given predefined articles (articles that have a rating associated with it already)
  - Article quality - featured, good, c-class, stub
  - Article importance - high, mid, low
- Learns proper phrases and wording
  - Destructive editing
  - Unreputable sources
  - Editor bias
- Classifies articles based on what it learns

Model Training

Load data...

```
Using TensorFlow backend.
Opening data...
Opening data/html...
        Reading kywW1B8qpW1N.html...
        Reading vNpA3FsQLhMR.html...
        Reading _NHr0O9MYULQ.html...
        Reading ryUagSdrzTpI.html...
        Reading ZlohaQwj7k8F.html...
        Reading u82JB2cUoJKY.html...
        Reading XA4-xty3zIkt.html...
        Reading 4btIEon1KWHe.html...
        Reading 6Q_qPKlMCAZp.html...
        Reading v4JNs_0Xq5xy.html...
        Reading EoHrzx5ffbBv.html...
        Reading NVA0Edfl_JlS.html...
```

Interpret data...

```
Building Model...
Compiling Model...
Training Model...
Train on 9435 samples, validate on 1665 samples
Epoch 1/100
9435/9435 [==============================] - 88s - loss: 0.2774 - val_loss: 0.2468
Epoch 2/100
9435/9435 [==============================] - 87s - loss: 0.2411 - val_loss: 0.2368
Epoch 3/100
9435/9435 [==============================] - 89s - loss: 0.2316 - val_loss: 0.2281
Epoch 4/100
9435/9435 [==============================] - 88s - loss: 0.2214 - val_loss: 0.2171
Epoch 5/100
9435/9435 [==============================] - 88s - loss: 0.2074 - val_loss: 0.2006
Epoch 6/100
9435/9435 [==============================] - 91s - loss: 0.1825 - val_loss: 0.1659
Epoch 7/100
9435/9435 [==============================] - 90s - loss: 0.1306 - val_loss: 0.1074
Epoch 8/100
9435/9435 [==============================] - 90s - loss: 0.0930 - val_loss: 0.0896
Epoch 9/100
```

Results

```
<li itemprop="name">
<a itemprop="url" target="_blank" data-mobile="false" href="http://abc11.com/" name="lpos=nav[header_desktop_wirestory]&lid=seos=nav[header_desktop_wire
story]&lid=section[U.S.]">
U.S.
</a>
</li>
<li class="http:abcnews.go.cominternational" itemprop="name">
<a itemprop="url" data-tab="true" href="http://abcnews.go.com/Internatem-info-wrap"><h1> <a class="realStory" name="lpos=newsfeed[story_37905262]&lid=[h
eadline]" href="/Entertainment/star-wars-force-awakens-deleted-scenes/story?id=37905262" data-redirect="false">Check nd burned off. None of the affected
 cars carried hazardous materials and none caught fire. Crews from 15 fire departments were on the scene. Canadian Pacific says the crash derailed seven
 empty cars a-old Gary Hank Thompson  dubbed the "bogus beggar"  pleaded guilty in U.S. District Court in Bowling Green to making false statements and r
epresentations to the Social Security Administration. Thompsot allow the Pacers to take control until late in the first half when Turner scored eight st
raight points to start a 13-2 spurt. That made it 49-37. New Orleans couldn't get closer than four the rest ondition to get Social Security benefits. Lo
cal news outlets report that 33-year-old Gary Hank Thompson  dubbed the "bogus beggar"  pleaded guilty in U.S. District Court in Bowling Green to making
 falsfive points out of halftime against Oklahoma, forcing the No. 2 seed to call a quick timeout. The Sooners answered with a handful of baskets aided
by a quick transition game, but the Aggies appear mucmage-container=".feed-item-figure" /></picture></div></figure><div class="text-container"> <div cla
ss="item-info-wrap"><h1> <a class="realStory" name="lpos=newsfeed[story_37909246]&lid=[headline]" hreader_desktop_wirestory]&lid=section[Log In]">Log In
</a>
<a class="profile-logged-in" data-tab="true" style="display: none;" data-affiliatename="abcn" href="#" thref="#" data-behavior="profile" data-ro/div>
</li>
</ul><div class="article-copy">
<p itemprop="articleBody">
</p><p itemprop="articleBody">
Myles Turner scored 24 points and had a career-high 16 rebounds on his 20th birthday, and C.J. Milursday night and will probably be in the lineup Friday
, Mackanin said. Herrera last played on March 12. UP NEXT Phillies RHP Jeremy Hellickson, named Wednesday as the opening day starter, will pitch S"http:
//abcnews.go.com/Sports/wireStory/pacers-celebrate-turners-big-birthday-beating-pelicans-37916078" data-ob-template="abc"></div>
<div class="social-footer-wrapper">
<ul class="article-social footrobras has been moving closer to Rousseff's inner circle in recent weeks. A recent poll by the respected Datafolha agency
says 68 percent of Brazilians surveyed want to see lawmakers vote to impeach image]" href="/Technology/york-international-auto-shows-coolest-concept-car
s/story?id=37898788" data-redirect="false"></a><figure class="feed-item-figure"><div class="img-wrap">"<picture> <source datao halve his match Friday ag
ainst Justin Thomas to reach the round of 16. "When you're 3 up and you're striking the ball well on a windy day with a difficult golf course, it's diff
icult to come from bea</a> </h1><div class="news-feed_item-meta"> </div> </div> </div> </article> <article data-id="37909128" class="news-feed-item w-im
ages" data-index="4"> <a class="story-link" data-id="37909128" name="ue" href="http://abcnews.go.com/Topics" name="lpos=nav[header_desktop_wirestory]&li
d=section[Topics]">
Topics
</a>
</li>
<li class="http:jobs.abcnews.comajobslist" itemprop="name">
<a itemprop="url" ding to school and attendance has dropped. ICE spokesman Bryan Cox referred questions about Holmes' comments about the wake-up call to
 Corrections Corp. of America, which manages the facility where Aco+ 'js/chartbeat.js'); document.body.appendChild(e); }
```

Outline Going Forward….