**Who’s Quoting the Bible:   
A Model for Quantifying and Categorizing Biblical Quotes and Topics in Any Text**

An app that quantifies and categorizes biblical quotes and topics in any text: Feed it documents from Classic Literature, Publications, Tweets, Songs, etc. and it identifies and quantifies any content that is sourced from the Bible.

Build a custom corpus: The Bible

* KJV, ASV, NIV

Predictive model that IDs direct quotations or allusions to ngram phrases from Bible

* Easier to ID direct quotes
* Harder to ID allusions that use extra-biblical language
  + E.g., nativity, Trinity, transubstantiation etc. that are a part of Christian lingo, doctrine, theology but not terms found in English language Bibles

Corpora possibilities:

1. New Testament vs Old Testament
   1. Look at intertextuality: References within Bible to other parts of Bible
2. Milton
   1. Does he directly quote or are his allusions more poetically phrased so that 3 and 4 token ngrams do not directly match Bible?
   2. What Bible translation did he read? What language did he write in?
      1. Latin, KJV, Tyndale, Coverdale, Wycliffe, Geneva, Douay-Rhiems, Bishops
      2. <https://en.wikisource.org/wiki/Bible_(Tyndale)>
      3. <https://en.wikisource.org/wiki/Bible_(Coverdale)>
      4. <https://en.wikisource.org/wiki/Bible_(Wycliffe)>
      5. Biblehub.com username: bh733 password: qx358
3. Theological/religious articles
   1. Compare frequency of biblical quotes across various denominational publications (do some denoms quote Bible more?)
   2. Compare the passages quoted across same corpora (are there different topics/themes emphasized by diff denoms?)
4. Tweets (filtered on some religious hashtag)
   1. What is the frequency of Bible quotes in tweets
   2. Are they topical (peak in tandem with particular news) or evenly distributed over time
5. Song lyrics (corpora for comparison: rock, country, hiphop, CCM, hymns)
   1. Compare frequency of biblical quotes across music genres
   2. Compare the passages quoted across same corpora

**How can I obtain these corpora? (download existing, request, web scrape, harvest)**

1. New Testament, Old Testament
   1. Multiple translations
   2. <http://www.hackathon.bible/data.html>
      1. [**ABS BibleSearch API**](http://webtools.bible/api.html)**– American Bible Society (free)**  
         Access via XML or JSON requests served over HTTP using RESTful resources, **over 300 versions** in 275 languages available in the API.
      2. [**Free Bible Downloads**](http://biblehub.net/)**(biblehub.net)**  
         Free access to Interactive eBibles and PDFs in 4 formats: .epub, PDF, Excel database, full text. Plus, there's a "**Ten-in-One Bible Database**" in parallel for quick access and comparison, with the following 10 versions, in Excel or full text formats:
   3. http://simoncozens.github.io/open-source-bible-data/
2. Milton
   1. Complete works or work by work
3. Theological/religious articles
   1. Contact denom publishing offices
4. Tweets (filtered on some religious hashtag)
   1. Harvesting now
5. Song lyrics (corpora for comparison: rock, country, hiphop, CCM, hymns)
   1. Compare frequency of biblical quotes across music genres
   2. Compare the passages quoted across same corpora

<https://www.independent.co.uk/arts-entertainment/books/features/restoring-holy-order-is-the-king-james-bible-the-only-version-we-should-celebrate-2105869.html>

* [Susan Elkin](https://www.independent.co.uk/author/susan-elkin)  [@SusanElkinJourn](https://twitter.com/SusanElkinJourn)  Sunday 17 October 2010

“The linguist David Crystal set out to discover systematically and scientifically – rather than relying on casual impression – just how many expressions in English derive solely from the KJB, rather than from earlier versions, or from much older pre-existing proverbs and other turns of phrase.

The result, published in *Begat: The King James Bible and the English Language*, … just 257 expressions rooted entirely in the KJB. Although this is a higher number than come from any other source (Shakespeare is, of course, the runner up) …”

Crystal proves that, without doubt, the Bible, in all its forms, has had an enormous influence on the English language. It is very difficult, however, to pin much of this down to the KJB in particular

it was the **protestant Latin version that Milton knew**. He was Cromwell's Latin secretary. He didn't need an English translation. The only evidence that he read the KJB is the one plagiarised line in Paradise Lost: 'She gave me of the fruit and I did eat.'" He points out, in passing, that this monosyllabic line is a perfect iambic pentameter