



Quotations on STOKES

from Johnson's Dictionary

Connie Harper

Amy Larner Giroux, PhD

UCF Center for Humanities & Digital Research



https://johnsonsdictionaryonline.org

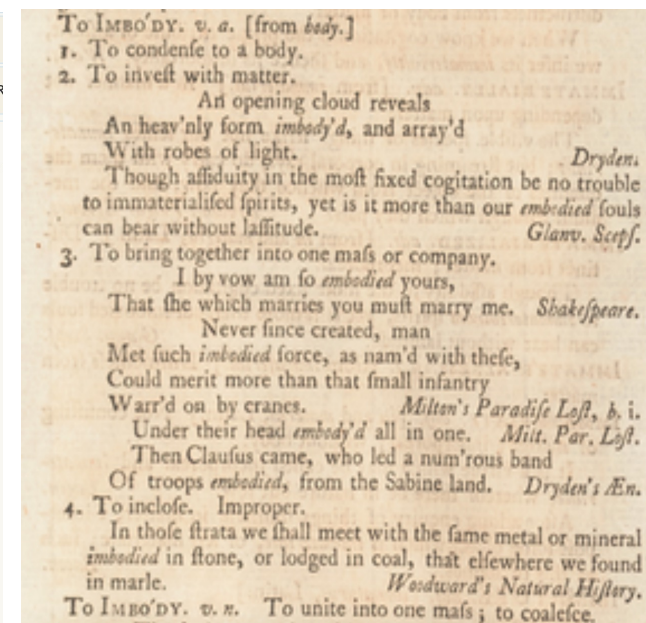
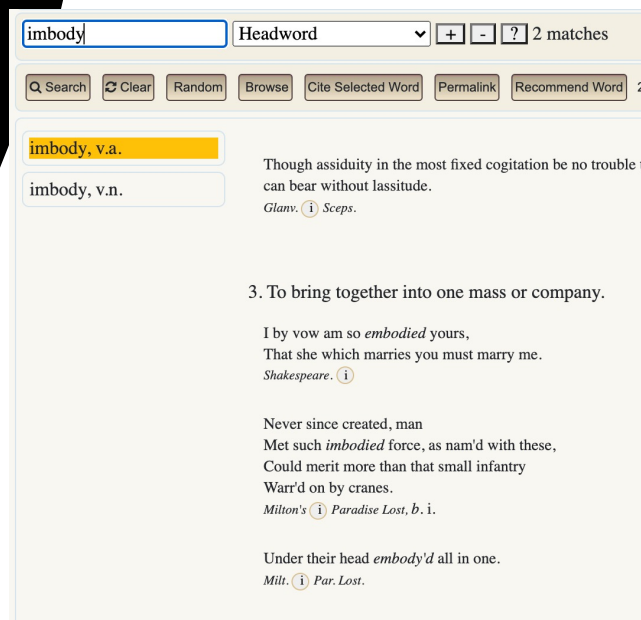
About the SJD

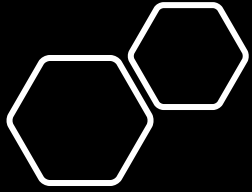
Johnson's dictionary was created in 1755 by Samuel Johnson and has been digitized by a group at UCF thanks to an NEH grant.

There are about 41,000 headwords and 120,000 quotations in the 1755 edition.

Johnson used quotes to show examples of word usage.

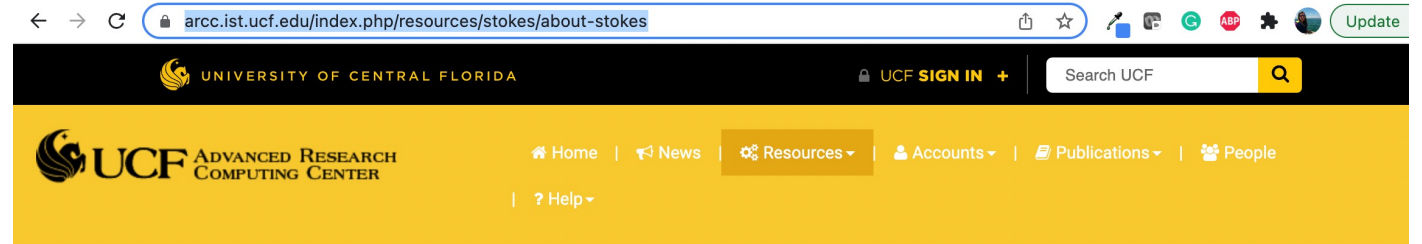
Some are paraphrased or misidentify the title. Some have the wrong attribution.





About STOKES

- UCF Advanced Research Computing Center
- 4848 compute cores
- Currently it takes 5 days to process 10,000 quotes on our server. We have 300,000 quotes in two editions.



About Stokes

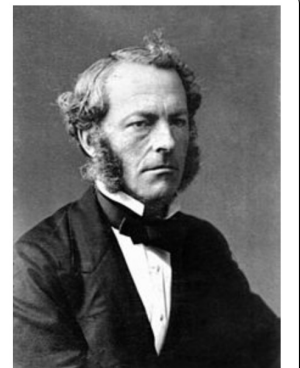
The Stokes HPC is named in honor of Sir George Gabriel Stokes who, through his wide range of contributions to mathematics, engineering, and natural philosophy, embodies the spirit of multidisciplinary, multiscale, and collaborative research across a wide range of disciplines.

Currently, the Stokes cluster consists of:

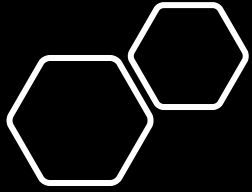
- 4,848 compute cores (Intel Xeon 64-bit processors)
- 56Gb Infiniband interconnect

In addition, the UCF ARCC also provides:

- 240TB of configured storage
- 20Gb connection to UCF campus core network
- 250kW power with an uninterruptible power supply and generator
- 100 tons of dedicated cooling



Sir George Gabriel Stokes
(1819-1903)



The Data and the Algorithm

An earlier team created algorithms using python to download the corpus text and to fuzzy match the quotes. The algorithm is written in python and is on our chdr server.

Ubuntu, MySQL, python3

From mortal or immortal minds Thus they
Breathing united force with fixed thought
Mov'd on in silence to soft Pipes that charm'd
Thir painful steps o're the burnt soyle; and now
Advanc't in view they stand, a horrid Front
Of dreadful length and dazzling Arms, in guise
Of Warriors old with order'd Spear and Shield,
Awaiting what command thir mighty Chief
Had to impose: He through the armed Files
Darts his experienc't eye, and soon traverse
The whole Battalion views, thir order due,
Thir visages and stature as of Gods,
Thir number last he summs. And now his heart
Distends with pride, and hardning in his strength
Glories: For never since created man,
Met such imbodyed force, as nam'd with these
Could merit more then that small infantry
Warr'd on by Cranes: though all the Giant brood
Of Phlegra with th' Heroic Race were joynd
That fought at Thebes and Ilium, on each side
Mixt with auxilliary Gods; and what resounds
In Fable or Romance of Uthers Son
Begirt with British and Armoric Knights;
And all who since, Baptiz'd or Infidel
Jousted in Aspramont or Montalban,
, or Marocco, or Trebisonde,
Or whom Biserta sent from Afric shore
When Charlemain with all his Peerage fell
By Fontarabbia. Thus far these beyond
Compare of mortal prowess, yet observ'd
Thir dread Commander: he above the rest
In shape and gesture proudly eminent

imbody | Headword | + | - | ? | 2 matches

Q Search | Clear | Random | Browse | Cite Selected Word | Permalink | Recommend Word | 2 Results Combined

imbody, v.a. | Thoug assiduity in the most fixed cogitation be no trouble to immaterialised spirits, yet is it more than our *embodied* souls can bear without lassitude.

imbody, v.n. | *Glav.* [i](#) *Seeps.*

3. To bring together into one mass or company.

I by vow am so *embodied* yours,
That she which marries you must marry me.
Shakespeare. [i](#)

Never since created, man
Met such *imbodyed* force, as nam'd with these,
Could merit more than that small infantry
Warr'd on by cranes.
Milton's [i](#) *Paradise Lost, b. i.*

Under their head *embody'd* all in one.
Mil. [i](#) *Par. Lost.*

Then Clausus came, who led a num'rous band
Of troops *embodied*, from the Sabine land.
Dryden's [i](#) *Æn.*

score	headword	author	author_work_author	work_filepath	work_url	etsc	content
63.3166	To Appro've.	Rogers.	A godly and learned exposition of Christs Sermon in the Mount: preached in Cambridge by that reuerend and iudicious diuine M. William Perkins. Published at the request of his exequutors by Th. Pierson preacher of Gods word. Whereunto is adioyned a twofol	Perkins, William, 1558-1602.	ebbo_texts/ebbo442A_god.txt	A09432.P4.xml	The first care and concern must be to approve himself to God by righteousness, holiness, and purity.
91.3208	To Imbo'dy.	Milton's	Paradise lost a poem written in ten books / by John Milton.	Milton, John, 1608-1674.	ebbo_texts/ebbo11230Parad.t	A50919.P4.xml	Never since created, man Met such imbodyed force, as nam'd with these, Could merit more than that small infantry Warr'd on by cranes.
57.1429	To l'ssue.	Dryden.	A commentarie or exposition, vpon the five first chapters of the Epistle to the Galatians: penned by the godly, learned, and iudiciall diuine, Mr. W. Perkins. Now published for the benefit of the Church, and continued with a supplement vpon the sixt chap	Perkins, William, 1558-1602.	ebbo_texts/ebbo460A_com.txt	A09383.P4.xml	A buzzing noise of bees his ears alarms; Straight issue through the sides assembling swarms.
56.3953	Adry'.		The wonders of the little world, or, A general history of man in six books : wherein by many thousands of examples is shewed what man hath been from the first ages of the world to these times, in respect of his body, senses, passions, affections, his vir	Wanley, Nathaniel, 1634-1680.	ebbo_texts/ebbo12008The_w_	A67489.P4.xml	He never told any of them, that he was his humble servant, but his well-wisher; and would rather be thought a malecontent, than drink the kingâ€™s health when he was not adry.
56.3953	Adry'.		The wonders of the little world, or, A general history of man in six books : wherein by many thousands of examples is shewed what man hath been from the first ages of the world to these times, in respect of his body, senses, passions, affections, his vir	Wanley, Nathaniel, 1634-1680.	ebbo_texts/ebbo12008The_w_	A67489.P4.xml	He never told any of them, that he was his humble servant, but his well-wisher; and would rather be thought a malecontent, than drink the kingâ€™s health when he was not adry.

Why do we need phase 2?

Nice!

1. Fuzzy matching algorithm works on the CHDR server.
2. Corpus is scraped and downloaded from sources.
3. Creates a spreadsheet with top 5 matches for each quote.

Better!

1. Currently it takes about 5 days to process 10,000 quotes. We have 300,000 quotes in 2 editions. The python program is written so that it can run on multiple cores. We think it can be adapted to run on STOKES without changing the core logic.
2. We will try to use the paywalled sources that the UCF Library has access to and limit our corpus to the two most reliable.
3. Make the output more specific with URL, title and catalogue number from corpus. I.e., Currently it might say “Dryden’s selections” but we would like to be a specific title where possible.
4. We would like our users to be able to search a canonical version of the quote from our data sources even if Johnson’s quote is incorrect.

Requirements – Phase 2

1. Work with the online corpora EEBO and ECCO and determine whether you'll need to have a download process. We will work with the sources to see if we can get the paywalled version of the data to work with.
2. Modify the existing python fuzzy search algorithm to run on STOKES. This will probably entail changing it to work with JSON data instead of querying the database.
3. Output a spreadsheet that has the input data plus the specific title of the work you found it in, a link to the Library of Congress's linked open data for the author, a link to the online text in which the quote was found, the match score, and the actual text of the quote you found.
4. Give us a master list of quotes, titles, and authors that Johnson used in his dictionary. This is sort of a fuzzy match within the dictionary itself. We need to be able to search a canonical quote and see where it is used in the dictionary even when it is abbreviated or misattributed, etc.



Questions?

Connie.Harper@ucf.edu

