XSEDE’s Text Analysis Gateway
and HPC support for humanists everywhere

Quinn Dombrowski
DH in Research IT group
Why would humanists use HPC?

- 3D modeling
- OCR
- Text analysis
3D modeling
friendship between rivals, and beauty taught the beholders chastity
on with his praises, but St. Stephen by him stay, and looks; and so they both perceived
a thing which floated drawing nearer and nearer to the banke, but rather by the fa-
vourable working of the Sea, then by any false industrious. They doubted a while what
it should be, till it was cast up even hard before them: at which time they fully saw
that it was a man. Whereupon running for pitie sake unto him, they found his hands
(as it should appear, constant friends to his life than his memory) fast gripping
upon the edge of a square small censer, which lay all under his breast; else in himselfe
no shew of life, so as the board seemed to be but a beere to carrie him alond to his
10 Sepulcher. So drew they vp a yong man of so goodly shape, and well pleasing fauour,
that one would thinke death had in him a lonely countenance; and, that though he
were naked, nakedness was to him an apparell. That sight increased their compassio-

15 n, and their compassion called vp their care; so that lifting his feete above his
head, making a great deale of stale water come out of his mouth, they layd him vp
some of their garments, and fell to rub and chafe him, till they brought him to to recov-
er both breath the seruant, and warmeth the companion of living. At length open-
ing his eyes, he gave a great groan; a dolorous note, but a pleasant dittie: for by that,
XSEDE Text Analytics Gateway

Mike Black1 and Drew Schmidt2
1University of Illinois, 2University of Tennessee

Background

- Humanities and social sciences historically dominated by qualitative methods.
- Researchers in these fields increasingly turning to quantitative methods and computational resources to answer larger questions.
- Still much resistance, often driven by a "technological familiarity gap".

So why expect users to start with this?

- Users often lack the background and have little access to training.
- No institutional rewards for developing technical skills.
- Programming vs GUI: recall vs. recognition.

TAG: The Text Analytics Gateway

Basic Text Analysis Tools Without the Programming

- Written mostly in R.
- Extensive re-use of existing R packages.
- UI served via shiny (3) and shiny server.
- Some custom high-performance C.
- Help files written in markdown, rendered on the fly.

Offer full analysis pipeline, by combining:
- Data retrieval
- Preprocessing
- Data retrieval
- Preprocessing into a single, complete workflow.

TAG Features

Detailed summary statistics

Includes LDAvis [2] for powerful topic modeling visualization:

And more...
- Processing: Text filtering and transformations.
- ViS: Dispersion plots and wordclouds.
- Analytic: LDA and n-gram modeling.
- Reproducibility: State management, script generation.

Timeline and Status

- November 12, 2014: Allocation awarded.
- April 6, 2015: PI connected to ECSS staff.
- April 6—March 15, 2015: Research technology stack.
- March 20, 2015: Development begins.
- July 7, 2015: Gateway stood up on ANS: Usability studies begin.
- Soon: live on Comet, tutorials, workshops.

Future Work

- Stand up gateway on Comet.
- Incorporate multidocument support.
- Add sentiment analysis feature.
- Develop usecase and community.

References


Acknowledgements

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575 and the XSEDE Science Gateways Program.

https://github.com/XSEDEScienceGateways/TAG
It’s been built,
but will they come?
Visual Text Explorer (University of Chicago)
VII. THE ADVENTURE OF THE BLUE CARBUNCLE I had...

VII. THE ADVENTURE OF THE BLUE CARBUNCLE I had called upon my friend Sherlock Holmes upon the second morning after Christmas, with the intention of offering him the compliments of the season. He was lounging upon the sofa in a purple dressing-gown, a pipe-nick within his reach upon the right, and a pile of crumpled morning papers, evidently newly studied, near at hand. Beside the couch was a wooden chair, and on

- holmes
- man
- that
- man's
- bird
- stone
- Baker
- said
- man
- hat
- goose
Next steps

● Gathering feedback on Text Analysis Gateway
● Presentation / workshop(?) at international digital humanities conference
● Mellon grant for planning project on broadening HPC support?
● Connecting up with other computation facilitator groups?