JISC DiD 3
Commonplace Cultures

Min Chen
Professor of Scientific Visualization
Chicago Members of the Project

- 2 postgraduate students
  - TBA
- Professor Robert Morrissey
  - PI, digital humanities, history
- Professor Ian Foster
  - CI, computational scientist
- Dr. Glenn Roe
  - Coordinator, text mining, French literature
- Dr. Mark Olsen
  - digital humanities, history

Oxford Members of the Project

- Dr. Alfie Abdul-Rahman
  - RA, visualization
- Professor Min Chen
  - PI, visualization
- Professor Nicholas Cronk
  - CI, French literature
- Dr. Glenn Roe
  - Coordinator, text mining, French literature
Similarity measures, and text alignment

- Chicago: ARTEL/PAIR and Ency./PAIR projects since 2009
- PhiloLogic, PhiloLine: English, French, Greek, Latin.

Four Levels of Visualization

1. **Disseminative Level (This is !)**
   - *A presentational aid for disseminating information or insight to others.*
   - *The creator does not expect to gain much new knowledge.*

2. **Operational Level (What?)**
   - *An operational aid that enables intuitive and/or speedily observation of captured data. Often part of routine operations.*
   - *Confirmatory observation, anomaly detection., etc.*

3. **Analytical Level (Why?)**
   - *An investigative aid for examining and understanding complex relationships (e.g., correlation, causality, contradiction).*
   - *Evaluating hypotheses, models, methods, algorithms and systems.*

4. **Inventive Level (How?)**
   - *A developmental aid for improving existing models, methods, algorithms and systems, as well as the creation of new ones.*
Four Levels of Visualization

1. **Disseminative Level (This is !)**
   - A presentational aid for disseminating information or insight to others.
   - The creator does not expect to gain much new knowledge.

2. **Operational Level (What?)**
   - An operational aid that enables intuitive and/or speedily observation of captured data (in routine operations, for external memorization)
   - Confirmatory observation, anomaly detection., etc.

3. **Analytical Level (Why?)**
   - An investigative aid for examining and understanding complex relationships (e.g., correlation, causality, contradiction).
   - Evaluating hypotheses, models, methods, algorithms and systems.

4. **Inventive Level (How?)**
   - A developmental aid for improving existing models, methods, algorithms and systems, as well as the creation of new ones.
Poetry Visualization (with Utah)

- Supporting close reading

Abdul-Rahman, et al., CGF, 2013
Four Levels of Visualization

1. **Disseminative Level (This is !)**
   - A presentational aid for disseminating information or insight to others.
   - The creator does not expect to gain much new knowledge.

2. **Operational Level (What?)**
   - An operational aid that enables intuitive and/or speedily observation of captured data. Often part of routine operations.
   - Confirmatory observation, anomaly detection., etc.

3. **Analytical Level (Why?)**
   - An investigative aid for examining and understanding complex relationships (e.g., correlation, causality, contradiction).
   - Evaluating hypotheses, models, methods, algorithms and systems.

4. **Inventive Level (How?)**
   - A developmental aid for improving existing models, methods, algorithms and systems, as well as the creation of new ones.
**Existing Work**

- **Text mining, e.g.,**
  - Filtering, lemmatization and stemming
  - Index term selection
  - Vector space model
  - Linguistic preprocessing
  - Sequence alignment
  - Clustering and classification
  - Summarization and storytelling

- **Visual analysis, e.g.,**
  - Detecting plagiarism [Ribler & Abrams, 2000]
  - Patent search [Koch et al., 2011]
  - Story evolution [Rose et al., 2009]
  - Document classification [Oesterling et al., 2010]
  - Text alignment [Kim, et al., 2011]
Document Readability Analysis

- Transform textual units to multivariate records
- Some computational analysis

Oelke, D., Spretke, D., Stoffel, A., Keim, D., Visual readability analysis: How to make your writings easier to read, IEEE VAST, 2010
Figure 8: Visual Analysis of 8 election agendas from the elections of the German parliament in 2009.

(a) Average Readability Score
(b) Feature: Vocabulary Difficulty
(c) Feature: Word Length
(d) Feature: Sentence Structure Complexity
Four Levels of Visualization

1. Disseminative Level (This is !)
   - A presentational aid for disseminating information or insight to others.
   - The creator does not expect to gain much new knowledge.

2. Operational Level (What?)
   - An operational aid that enables intuitive and/or speedily observation of captured data. Often part of routine operations.
   - Confirmatory observation, anomaly detection., etc.

3. Analytical Level (Why?)
   - An investigative aid for examining and understanding complex relationships (e.g., correlation, causality, contradiction).
   - Evaluating hypotheses, models, methods, algorithms and systems.

4. Inventive Level (How?)
   - A developmental aid for improving existing models, methods, algorithms and systems, as well as the creation of new ones.
Four Assertions

- **statistical methods** alone cannot
  - convey an adequate amount of information for humans to make informed decisions

- **algorithms** alone cannot
  - encode an adequate amount of human knowledge about relevant concepts, facts, and contexts

- **visualization** alone cannot
  - effectively manage levels of details about the data or prioritize different information in the data

- **direct interaction** with data alone
  - isn’t scalable to the amount of data available

hence the need for
- visualization
- interaction
- analysis & interaction
- analysis & visualization
Visual Analytics for Supporting Sketch-based Video Search

Legg, et al., IEEE TVCG, 2013
System Architecture and Components

Digital Humanity Scholar

- Database Management
- Commonplaces Visual Analytics
- Multi-rule Text Mining for Commonplaces
- Rule Ensemble & Rule Space Visual Analytics

Updates and synchronisation management

- Offline Database of Commonplaces
- Offline Text Corpora
- Supporting Ontologies
- Rule Collection

Online Database of Commonplaces

- User Space
- Visualization Gallery
- Interactive Visualization
- Database Search
- User-centred Error Report & Rule Suggestion

Error reports

Suggested rules

Web interface

Ordinary Users
Scheduled Activities

- **Appointment**
  - *Oxford*: Dr. Alfie Abdul-Rahman
  - *Chicago*: two postgraduates

- **Data access**
  - *Chicago*: signing the agreement

- **Sharing references**
  - *Dropbox*

- **Visits**
  - *May*: Chen visits Chicago
  - *July*: Morrisey & Roe visit Oxford
  - *October*: Chen visits Chicago

- **Software development**
  - *May-June*: requirement analysis and design
  - *July-October*: development
  - *October*: testing and evaluation

- **Publications**
  - *December*: submitting paper(s)
Acknowledgement

University of Oxford
- Alfie Abdul-Rahman (PhD, 1989-1992)
- Saiful Khan
- Eamonn Maguire
- Simon Walton
- Hui Fang

- Colleagues in OeRC, OCCAM, ...

Swansea
- Rita Borgo
- Phil W. Grant
- Ian Griffiths
- Mark W. Jones
- Bob Laramée
- Adrian Morris
- Tavi Murray
- Irene Reppa
- Kilian Scharrer
- Ian Thornton
- ROs and PhDs (below)

Stuttgart
- Tom Ertl
- Daniel Weiskopf
- Ralf Botchen ...

Rutgers
- Deborah Silver
- Carlos Correa

Purdue (VACCINE)
- David Ebert

Heidelberg
- Heike Jänicke

Utah
- Chris Johnson, Kate Coles, Julie Lein, Miriah Meyer
- Chuck Hansen

Cardiff
- Andrew Aubrey
- Dave Marshall
- Paul Rosin
- Gary Tam

RIVIC
- Nigel John
- Ralph Martin
- Reyer Zwiggelaar

Past PhDs and ROs:

- C.-Y. Wang (PhD, 1989-1992)
- Abdula Haji Tabil (PhD, 1990-1994)
- Mike Bews (PhD, 1992-1996)
- Malcom Price (MPhil, 1997-1998)
- Adrain Leu (PhD, 1996-1999)
- Simon Michael (PhD, 1996-1999)
- Steve Treavett (PhD, 1997-2000)
- Mark Kiddell (RA, 1999-2001)
- Ben Smith (TCA, 1999-2001)
- S.-S. Hong (PhD, 1998-2002)
- Abdul Haji-Ismail (PhD, 1998-2002)
- Andrew S. Winter (PhD, 1999-2002)
- David Rogeman (PhD, 1999-2003)
- Paul Adams (TCA, 2002-2004)
- Tim Lewis (RA, 2004-2005)
- Gareth Daniel (PhD, 2001-2004)
- David P. Clark (PhD, 2001-2005)
- Dave Bown (RA, 2005)
- Ann Smith (PhD, RA, 2001-2006)
- Siti Z. Zainal Abdin (PhD, 2003-2007)
- Alfie Abdul Rahman (PhD, RA, 2004-7)
- Joanna Gooch (PhD, 2004-2007)
- Shoukat Islam (PhD, RA, 2004-2009)
- David Chisnall (PhD, RA, 2005-2008)
- Phil Roberts (RA, 2005-2008)
- Rudy R. Hashim (PhD, 2005-2008)
- Dan Hubball (MPhil, 2007-2008)
- Owen Gilson (PhD, 2006-2009)
- Lindsey Clarke (PhD, 2007-2010)
- Heike Jänicke (RO, 2009-2010)
- Farhan Mohamed (PhD, 2008-)
- Ed Grundy (PhD, 2009-)
- Rita Borgo (2009-2011)
- Hui Fang (2009-2011)
- Yoann Drocourt (PhD, 2010-2011)
- Karl Proctor (PhD, 2009-2011)
- Andrew Ryan (PhD, 2010-2011)
- Phil Legg (RO, 2010-2011)
- David Chung (PhD, RA, 2010-2011)
- Matthew Parry (MPhil, RA, 2010-2011)
- Richard M. Jiang (RO, 2010-2011)
- Brian Buffy (RO, 2011, 2013)
- Kai Berger (RO, 2012-2013)
- Karl Proctor (RO, 2012-2013)
- Jeyan Thiyagalingam (RO, 2013)