REPURPOSING LIBRARY COLLECTIONS IN SUPPORT OF DIGITAL HUMANITIES RESEARCH
TRANSFORMATIVE TRAJECTORIES

• Libraries from Repositories to Workshops
• Librarians from Curators to Collaborators
• Individual works to Networked Collections
A continuum of practice
Inside and outside the library
Smaller-scale corpus building and analysis
News and press analysis projects
Network analysis projects
SUBJECT: CASE XYZ - UFO RESEARCH/ORD REQUEST FOR ADDITIONAL INFORMATION

REFERENCE: A. B.

1. ORD HAS EXHIBITED SOME INTEREST IN THE WORK OF A QUALIFIED ANALYST IS CURRENTLY ATTEMPTING TO EVALUATE SYSTEM AND HAS REQUESTED ADDITIONAL INFORMATION.

2. WE NOTE THAT IN SOME OF THE EARLIER CORRESPONDENCE FROM YOUR OFFICE (REFERENCE B) MENTION WAS MADE OF THE

REQUEST OF DOING SOME HOW ADDITIONAL INFORMATION...
The government originally held back the majority of memos that contained the word "Boulder". (Source: Matthew Connelly)
DECLASSIFICATION ENGINE: COMPARING REDACTED AND UNREDACTED VERSIONS

[Vietnam] An Assessment of the Enemy Sapper Threat [problem of...
Feb 11, 1954

1975

1985
VENDOR AND GOVERNMENT SOURCES
CHARTEX: DISCOVERING SPATIAL DESCRIPTIONS AND RELATIONSHIPS IN MEDIEVAL CHARTERS
WWW.CHA Tex.ORG
@CHA TexPROJECT
CHARTEX ARCHITECTURE

Charter documents → Natural language processing → Analysed individual documents → Data mining → Analysed integrated documents

ChartEx workbench
Grant by Thomas son of Josce goldsmith and citizen of York to his younger son Jeremy of half his land lying in length from Petergate at the churchyard of St. Peter to houses of the prebend of Ampleford and in breadth from Steyngate to land which mag. Simon de Evesham inhabited;
LAYERS OF NLP

- Semantic: Gluing phrases together using semantic expectations
- Phrasal: Building phrases participating in subcat and coord etc
- Syntax: Building local syntactic structures
- Lexical: Identify multi-word expressions that behave like words
- Token: Identify properties of individual words (known types etc)
- Word: Bootstrapping lexical access using external tools (POS etc)
- ELF: A lexical approach to multi-word/sentence processing
- DATR: A default inheritance language for defining lexicons
MATCHING RELATIONAL INFORMATION
CHARTEX SOURCES

• The Vicars Choral (University of York), 125 charters manually annotated, English, 5,000 charters (dated).

• Borthwick (Borthwick Institute, University of York), 55 charters manually annotated, English.

• DEEDS (University of Toronto), 49 charters manually annotated, Latin, over 10,000 charters.

• Wards2 (The National Archives, UK), 48 charters manually annotated, English, 7,000 charters.

• Cluny (University of Columbia), 50 charters manually annotated, Latin, over 5,000 charters (dated).
IN THE LIBRARY, OUTSIDE THE LIBRARY: WHY NECESSARILY IN THE LIBRARY?
IN THE LIBRARY, OUTSIDE THE LIBRARY: WHY NECESSARILY IN THE LIBRARY?

Access to resources
New generation of academic librarians
A new vision of librarianship
The new interdisciplinaria
A crossroads, a neutral player, a broker
Greater potential for reapplication of lessons and products
LIBRARIAN AS ADVISOR

About resources
About local expertise
About tools and methodologies
LIBRARIAN AS NEGOTIATOR

About Access:
- Pilot
- Pay
- Permission

About terms of use and protection of intellectual property

About venue
- API
- Vendor sandbox
- Delivered files
- Already own?
LIBRARIAN AS PARTICIPANT

Preparation of material
Consultant on project concept, design, interface, functionality
Grant writer
Inputter, annotator, coder
Assessment and evaluation
Disseminator and instructor
CHALLENGE FOR LIBRARIES

Developing new skills
Fostering new partnerships
Hiring new kinds of staff
Revising job descriptions
Carving out time
MORE INFORMATION


• “CHARTEX: Discovering Spatial and Temporal Descriptions and Relationships in Medieval Charters.” [Whitepaper, 12 February 2014]

• www.chartex.org

• www.declassification-engine.org