GOALS, DIRECTIONS, POSSIBILITIES, THOUGHTS

Fe6 Project

A: Quarterly goals (2017-4)

- Pipeline: push complex annotations through 2&3; check and adjust 4&5&6 components

- CMS: complete: book import, FROntIER integration, and non-COMET and test lines

- COMET: enhance (if needed) for production work

- FS-ingest: settle on an ingest target, get buy-in from FS, and adjust to meet FS input requirements

- ConstraintEnforcer: complete ancestor constraint-checker demo

- FROntIER: integrate into workflow

- GreenDDA: test relationships; code COMET training data generator

- GreenFIE: submit paper

- GreenQQ: write paper; create COMET-based interface

- ListReader:

- OntoES: design development interface and begin implementation

- OntoSoar: deploy on dithers

A: Ingest into FS

- FamilyTree (continue discussions, follow-up on Knox County temple submissions)

- Obituaries (got GedcomX specs, looks like it could work, follow-up if nothing better)

- Scanned books (Ty Davies & Jon Morrey discussion leading to discussions with Jonathan Cranford)

- LLS and CDS (reconcile requirements with FS for automatic ingest)

B: FS future projects (wish list, to continue to encourage and help realize):

- Highlighted, document-first output by search system

- HyKSS-like search engine over all content

- Patron ingest via COMET, Ddupe-like entity resolution, and semi-automatic FamilyTree update

- Green system of the whole – constant improvement while doing real work

Research

A: (me, from Thomas) follow up on TKDE paper (July 17 resubmission)

- Name/Place/Date-level text abstraction (w/ DL & SL)

- high-level nested pattern discovery in a second pass with record-level text abstraction

A: (Liddle) GreenFIE paper (DKE), Heinrich journal paper (emisa)

- CM-based document reading system supported by ensemble & PRF results (emisa experience report)

- extraction by layout for forms & diagrams (especially for Jaipu)

A: (Lonsdale)

- GreenDDA (w/ GN)

- preprocess with named-entity recognizer (does this give a significant boost?)

- long-term directions: anaphor resolution, XNL parser; declarative rule specification; rule learning

A: (Nagy) DAS paper (papers due: 20 Nov 2017)

- extend GreenQQ to work with our set of data

- create a COMET-based interface for GreenQQ

B: (Woodfield)

- quality assessment (of merge, of an individual’s genealogy, of FamilyTree)

- Bayesian reasoning for assessing quality

C: Grand Challenges

- “Green Interaction” (systems that improve while being used for real-world applications)

- “Teaching Computers to Read” (cognitive computing grand challenge)

- “Web of Knowledge” (WoK vision with FamilySearch as an example)