GOALS, DIRECTIONS, POSSIBILITIES, THOUGHTS

Fe6 Project

A: Quarterly goals (2018-2)

- FS-ingest: work toward enabling others (missionaries, students, …) to process a book

- CMS: design and implement workflow; complete book import and non-COMET lines

- Pipeline: push complex annotations through Phase 3; refactor Annotator2 I/O?

- COMET: enhance (if needed) for initial production work; rework?

- ConstraintEnforcer: enhance Tree Sweeper app; code merge checker

- FROntIER: extend ontology with Mother, Father, Parent?

- GreenDDA/GreenML: install new Stanford NLP tools, test relationships

- GreenFIE: submit paper

- GreenQQ: solidify interface design and begin coding

- ListReader:

- OntoES: do backend coding for CMS configuration and testing; implement user interface

- OntoSoar:

A: Ingest into FS

- Possibilities (?):

- FamilyTree (import via eventual human checked person info doc through D-Dupe interface)

- Scanned books (eventual HyKSS-like search)

- LLS and CDS (initially, not hopeful, but the “soon” to be LLS Tree looks promising)

- Jon Morrey’s suggestion: run 1,000 books through the system

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, D-Dupe-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 (Dec 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), Dagstuhl position paper

- domain-specific conceptual models (Dagstuhl seminar)

- extraction by layout for forms & diagrams

A: (Lonsdale)

- GreenDDA (w/ GN), GreenML & GreenOTS

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

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

A: (Nagy) GreenQQ journal paper (enabling technology: interface, integration into pipeline)

A: (Woodfield) ER’18 paper follow-up

- 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): EMISA paper

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