GOALS, DIRECTIONS, POSSIBILITIES, THOUGHTS

Fe6 Project

A: Quarterly goals (2018-3)

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

 - CMS: design and implement Green workflow for COMET (import, export, indexing)

 - Pipeline: complete Green workflow for initial import & export and for GreenQQ indexing

 - COMET: rework design and implementation for Green workflow

 - ConstraintEnforcer: enhance Family retraction/revision handler; design merge/unmerge checker

 - FROntIER:

 - GreenDDA/GreenML: test relationships; design/test for inclusion in the Green workflow

 - GreenFIE: submit paper

 - GreenQQ: solidify for use in Green workflow; implement Advanced User Interface

 - ListReader:

 - OntoES: reassess design for use in Green workflow indexing

 - OntoSoar:

A: Ingest into FS

 - Green workflow interface:

 - Import to search repository

 - HyKSS-like search over search repository

 - Export from search repository to COMET

 - Export from COMET to FamilyTree (with human checked person info via D-Dupe-like interface)

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

Research

A: (Liddle) GreenFIE paper (DKE)

 - Green System of the Whole based on GreenQQ patron interface (emisa experience paper, w/ GN)

 - 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

A: (Woodfield)

 - quality assessment (duplicate detection on import, merge/unmerge in FamilyTree)

 - Bayesian reasoning for assessing quality

B: (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

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)