**Fe6 Ensemble: Developer’s Guide**

**CMS** (**C**OMET **M**anagement **S**ystem)

* Use Chrome as your browser for CMS
* Navigate to <http://deg.byu.edu/comet2/home.php> (... eventually, comet, not comet2)
* Logon with an admin account
  + Establish an administrator account (…?? give out a special PIN for “M” (adMinistrator) accounts)
  + Establish a regular account (obtain PIN from us) and … ?? send email to request admin privileges
* The options on the admin page are the bold section titles below

**Import Book**

* Preparation for import (outside of CMS)
  + Scan book, creating a single pdf
  + Rename the pdf with a Short Title for the book written in camel-case with no spaces
  + Add OCR (Tools > Text Recognition > Aa In This File)
  + Split into pages (Tools > Pages > Split Document); no label, make “.” the separator
  + Put the page pdf’s in a directory by themselves
* Provide information for import
  + Browse for the pdf of the book
  + Browse for the directory of pdf pages
  + Enter the bibliographic information for the book
* Click upload (from here, we (1) select the next available number for the book, (2) update the database, (3) generate all directories, (4) create the meta.xml file, (5) put the meta.xml file and the pdf of the book in the book’s directory, (6) add the empty book-specific name- and place-authority files, (7) upload the pdf pages into 1.pages, and (8) run PDF\_INDEXER to create the html, png, txt, and xml files)

**Configure Extraction Tools** (See explanation below for each tool)

* For FROntIER, and currently for OntoES, development outside of CMS, and we must provide a way to upload the developed ontologies from within CMS; an upload with a browse should be sufficient for now; eventually, we want to also be able to test the extraction ontologies by letting the developer view the results in COMET

**Initiate Pipeline Run**

* Select book
* Select page range
* Choose tools to use (for now, only OntoES)
* Choose run type (only “without COMET” is currently fully functional)
* Click Run

**Obtain Results** (not a clickable admin option; results are at the end of the pipeline run)

* Obtain final txt and pdf files at http://deg.byu.edu/<book short title>/
* Download option: … ?? select book, then download all of them to a directory

**FROntIER (development on personal computer)**

* Install software
  + Install Java 8 Runtime Environment (<http://www.oracle.com/technetwork/java/javase/downloads/jre8-downloads-2133155.html>)
  + Mac users also need the Java SE Development Kit 8. (<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>)
  + Download Workbench jar (<http://deg.byu.edu/WorkbenchShell-1.0-jar-with-dependencies.jar>)
  + Download FROntIER jar (<http://deg.byu.edu/FROntIER-1.2-jar-with-dependencies.jar>)
* Establish working directories (see **OntoES** below)
* Set environment variable (see **OntoES** below)
* Download FROntIER’s base ontology (… link to base ontology)
* Develop extraction rules (see **OntoES** below)
* Test extraction rules (see **OntoES** below)
* Upload ontology from within CMS

**GreenFIE** (not yet available)

**GreenDDA** (not yet available)

**ListReader** (not yet available)

**OntoES** (development within CMS not yet available)

**OntoES (development on personal computer)**

* Download/Install software
  + Install Java 8 Runtime Environment (<http://www.oracle.com/technetwork/java/javase/downloads/jre8-downloads-2133155.html>)
  + Mac users also need the Java SE Development Kit 8. (<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>)
  + Download Workbench jar (<http://deg.byu.edu/WorkbenchShell-1.0-jar-with-dependencies.jar>)
  + Download OntoES jar (<http://deg.byu.edu/OntoES-1.0-jar-with-dependencies.jar>)
  + Download FROntIER jar (<http://deg.byu.edu/FROntIER-1.2-jar-with-dependencies.jar>)
* Establish working directories
  + In Documents create directories: dithers.byu.edu\data\fe6
  + In fe6 create directory: <Book #>.<Short Title> (same as in CMS)
  + In <Book #>.<Short Title> create two directories: 1.pages and 2.tools. (Note for Mac users: When a folder is named “1.pages,” it will show up as a word processing document. However, if you right click and select “show package contents,” it will open as a folder and run normally for the software. Do this and copy-paste the pages of the book into this folder. Do not try to open it as a normal folder, as the OS will try to open it as a pages document and break.)
  + Download the 1.pages directory from CMS
  + In 2.tools create directory: OntoES
  + In OntoES create three directories: Person, Couple, and Family
* Set environment variable
  + Windows
    - Control Panel > System and Security > System > Advanced system settings > Advanced tab > Environment Variables
    - Under System variables click on New
      * For Variable name, type in FE6\_ROOT\_PATH
      * For Variable value, copy&paste in the path to your fe6 directory (e.g. C:\Users\embley\Documents\dithers.byu.edu\data\fe6)
  + Mac
    - Open a terminal
    - Make sure you are at your initial directory (i.e. /Users/*username*)
    - To access the environment variable, type “nano .bash\_profile”. This will access the file containing all of your current environment variables.
    - Create a comment to know what the environment variable is (e.g. “# DEG genealogy path”). Underneath that line, type “export FE6\_ROOT\_PATH=“ followed by the directory of the fe6 folder. (e.g. “export FE6\_ROOT\_PATH=“/Users/*username*/documents/[dithers.byu.edu/data/fe6](http://dithers.byu.edu/data/fe6)")
    - Use the command control-o (“o” as in otter) and then hit enter to make sure that it is saved. Then the command control-x exits the .bash\_profile.
    - Because the environment variable is not initialized in the current terminal, close the existing terminal and open a new one.
* Download base ontologies into the OntoES directory
  + PersonExtrOntology.xml (… link to download from the dithers common directory)
  + CoupleExtrOntology.xml (… link to download)
  + FamilyExtrOntology.xml (… link to download)
* Develop ontology-snippet extraction rules
  + Open the Workbench
    - Windows: double-click on Workbench jar
    - Mac: open a terminal and change directory to where the Workbench is located. Then, type into the terminal “java -jar WorkbenchShell-1.0-jar-with-dependencies.jar”
  + Open a base ontology
    - Click on the open icon
    - Browse to the OntoES directory and double-click on an ontology
  + Open ontology-snippet tool (Tools > Show Ontology Snippet Editor)
    - Click on add new Ontology Snippet and then on the New Ontology Snippet that pops up
    - Change the Name to whatever you wish to call the rule
    - Develop a regular expression for the extraction rule
      * Open a regular-expression development tool on the web (e.g. <http://www.regexr.com/>)
      * Put one of the txt files from 1.pages into the tool and develop a regular expression with capture groups to extract what you want for the rule
      * Copy the regular expression and paste it into the Ontology Snippet Expression text box
      * Add Predicate Mappings
        + Click on “+” for the rule you are developing and then on Predicate Mappings
        + Add a rule and then click on object and relationship names to add mappings for capture groups
* Test extraction rules
  + Execute the extraction ontology
    - Open a command prompt and cd to wherever you put your jar files (I suggest the desktop, thus cd \Users\embley\Desktop)
    - Execute: java -cp OntoES-1.0-jar-with-dependencies.jar edu.byu.deg.extract.OntoESrunner -book %BOOK% -range %START\_PAGE%-%END\_PAGE% (where %BOOK% is <Book #>.<Short Title>, e.g. 000001.TheElyAncestry, and %START\_PAGE%-%END\_PAGE% is a page range, e.g. 573-573 for just a single page, i.e. java -cp OntoES-1.0-jar-with-dependencies.jar edu.byu.deg.extract.OntoESrunner -book 000001.TheElyAncestry -range 573-573)
  + Check the results
    - Results will be in the Person, Couple, and Family directories under OntoES
    - In the Workbench, browse to and double-click on a result file.
    - To see the results right-click on a relationship connector (e.g. the center of the line connecting the Person and the Name object sets) and then on Edit Relationships
* Upload ontologies from within CMS

**OntoSoar** (not yet available)