MetatMap Migration to Lexical Tools Java APIs – Functional Requirements

Test log	Functional Description	Requirements: Algorithm & Lexical Tools APIs	Test (word)	Test (base) (419,749)	Test (spVar) (529,820)
test_bff.log	• Retrieve all uninflected (base) forms, including citation forms and its spelling variants • Not case sensitive (of the input) • Include input term is not in LEXICON	• If it is in Lexicon: ▪ Retrieve all base forms (-f:b) ▪ Retrieve all spelling variants of the retrieved base forms with category and inflection (-f:s) ▪ Remove duplicated results ▪ Convert results to ASCII • else: ▪ Retrieve all base forms by rules (-f:b) ? ▪ No spelling variants (since it is not known by LEXICON) • lvg -f:b:s • toAscii	• Candomble • Styrenation • Munoz-gonzalez • Spiculopteragia bohmi • Saw • AAAS • Bacteria • manikin • sheep • cattle • mankind • SawSaw	• 145 (from 774) • All 145 (non-Ascii) • Eg. Boecking	• 387 • 260 (ASCII: not in LEXICON) • 7 (ASCII: new relationship) • 120 (problems in C inflectional morphology)
test_cff.log	• Retrieve citation form	• Retrieve citation form • Convert result to ASCII • Return null if no citation found • lvg -f:Ct • toAscii	• mum (E0041164) • muTBS (E0659182	• 32 (from 304) • 26 (ASCII: not known to LEXICON) • 4 (ASCII: data) • 2 (ASCII: C bug)	• 267 • 260 (ASCII: not in Lexicon) • 7 (ASCII: new relationship)
test_dmv.log	• Retrieve derivational variants • Not case sensitive of input • Use no restriction rules (-kd:3) • Apply min. term length (default = 3) • Apply min. stem length (default = 3)	• If it is in Lexicon (-f:Ln): ▪ Retrieve all categories and inflections (-f:Ln) ▪ Retrieve all derivational variants with category and inflection (-f:d) ▪ Remove duplicated results ▪ Convert results to ASCII • else: ▪ Retrieve all possible categories and inflections (-f:L) ▪ Follow above steps after step-1	• zonated • zap • Eastwards • Zoic • Red • Rist • SawSaw	• 732 (from 419,502) • 732 (difference of derivational facts) • Expect 732 to go higher after fixing bugs (several	• TBD

	• Include input term is not in LEXICON	• lvg -f:Ln:d –kd:3 (in Lexicon) • lvg –f:L:d –kd:3 (not in Lexicon) • ToAscii		thousands)	
test_dmvR.log	• Retrieve derivational variants that are known to LEXICON • Include input term is not in LEXICON	• lvg -f:Ln:d –kd:3 • ToAscii • lvg –f:E	• see dmv	• TBD	• TBD
test_lfc.log	• Retrieve categories • Include input term is not in LEXICON	• If it is in Lexicon (-f:Ln): ▪ Retrieve categories from LEXICON (-f:Ln) • else: ▪ Retrieve possible categories from rules (-f:L)	• See bff	• 26 to 26 • All 26 (non-ASCII)	• 260 • All 260 (non-ASCII)
test_vcf.log	??	• lvg -f:i:s • lvg –f:Ct • toAscii	•	• TBD	• TBD
test_vf.log	• Retrieve all inflectional variants, including spelling variants. • Include input term is not in LEXICON	• If it is in Lexicon (-f:Ln): ▪ Retrieve all spelling variant with category and inflection (-f:s) ▪ Retrieve all inflectional variants of the retrieved spelling variants with category and inflection (-f:i) ▪ Use restriction option 2 (LEXICON then rules, -ki:2) ▪ Convert inflections to simple inflections ▪ Mark spelling variants ▪ Remove duplicated results ▪ Convert results to ASCII • else: ▪ Retrieve inflectional variants (-f:i) ▪ Follow above steps after step-2. • lvg -f:i:s or –f:s:i • toAscii	• See bff	• 40248 to 126,823 (bug fixed)	• TBD
c_lex_from_input	??	??	• TBD	• TBD	• TBD