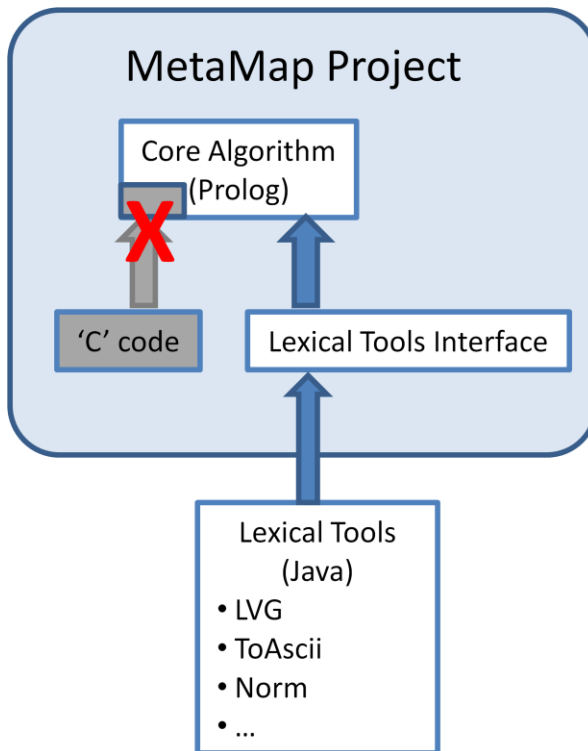


Introduction: Functionality of Lexical Tools Interface for MetaMap

1. Software Component



2. Life-Cycle

- *Requirements (functional specification)*
 - Report 1 by Francois : *Migrating MetaMap/SemRep to LexAccess* (March)
 - Report 2 by Francois : *Status of lexAccess Functionality for MetaMap* (Sept.)
 - From C-code results?
 - MetaMap documents, published paper, etc.?
- *Analysis, Design, Implementation*
 - Find equivalent Lexical Tools APIs (commands) to all 10 functions on report 1
 - Prolog interface to Java code
 - Implement Java code, algorithm and Lexical Tools APIs on report 2
- *Testing*
 - Lexical Tools API test (on examples from reports)
 - Prolog interface functional test (on examples from reports)
 - MetaMap test
 - Match C-code results, fix C-code bug, and any found bugs
- *Build & Deliver*
 - Prolog Lexical Tools Interface
 - Match C-code
 - If not, provide explanation (such as bug in C-code, ASCII conversion)
- *Maintenance & Enhancement*
 - Upgrade to latest Lexical Tools annually

3. MetatMap Tests Status

Test log	Original C function	Requirements (Description)	Lexical Tools APIs	Status
test_bff.log	<ul style="list-style-type: none"> c_lex_form_input 	Retrieve known spelling variants of uninflected (base) form of the input	<ul style="list-style-type: none"> lvgl -f:b:s toAscii 	<ul style="list-style-type: none"> 774 to 145 All 145 (non-Ascii) Eg. Boecking Done
test_cff.log	<ul style="list-style-type: none"> c_lex_form 	Retrieve citation form	<ul style="list-style-type: none"> lvgl -f:Ct toAscii 	<ul style="list-style-type: none"> 304 to 32 All 145 (non-Ascii) Done
test_dmv.log	<ul style="list-style-type: none"> c_lex_form_cats c_dm_variants 	TBV	<ul style="list-style-type: none"> lvgl -f:d – kd:3 –cf:2 toAscii 	<ul style="list-style-type: none"> Complete Lexical tests Complete Interface test Work on MetaMap test
test_dmvR.log	<ul style="list-style-type: none"> c_lex_form_cats c_dm_variants c_lex_is_a_form 	TBV	<ul style="list-style-type: none"> lvgl -f:d – kd:3 –cf:2 ToAscii lvgl -f:E 	<ul style="list-style-type: none"> Complete Lexical tests Complete Interface test Work on MetaMap test
test_lfc.log	<ul style="list-style-type: none"> c_lex_form_cats 	Retrieve category	<ul style="list-style-type: none"> lvgl -f:Ln 	<ul style="list-style-type: none"> 26 to 26 All 26 (non-Ascii) Done
test_vcf.log	<ul style="list-style-type: none"> c_lex_cit c_lex_form c_get_varlist 	TBV	<ul style="list-style-type: none"> lvgl -f:Ct lvgl -f:i:s toAscii 	<ul style="list-style-type: none"> Complete Lexical tests Complete Interface test Bug for uncount noun Work on MetaMap test
test_vf.log	<ul style="list-style-type: none"> c_lex_form c_get_varlist 	TBV	<ul style="list-style-type: none"> lvgl -f:i:s toAscii 	<ul style="list-style-type: none"> Complete Lexical tests Complete Interface test Bug for uncount noun Work on MetaMap test
??	??	Approximate match	??	<ul style="list-style-type: none"> TBD

MetatMap Tests Status

Test log	Requirements (Description)	Lexical Tools APIs	Test (base) (419,749)	Test (sp) (529,820)	Notes
test_bff.log	Retrieve known spelling variants of uninflected (base) form of the input	<ul style="list-style-type: none"> • lvg -f:b:s • toAscii 	<ul style="list-style-type: none"> • 774 to 145 • All 145 (non-Ascii) • Eg. Boecking 	<ul style="list-style-type: none"> • 387 • 260 (non-ascii) • 127 bugs fixed & features enhancements 	<ul style="list-style-type: none"> • Fact in C only include irreg (222,768 vs. 1,488,144) • Fact in Java (case insensitive), Eg. Boss • case bug, Eg. AAAS • Ascii issue
test_cff.log	Retrieve citation form	<ul style="list-style-type: none"> • lvg -f:Ct • toAscii 	<ul style="list-style-type: none"> • 304 to 32 • All 32 (non-Ascii) 	<ul style="list-style-type: none"> • 267 • 260 (non-ascii) • 7 other ascii conversion issue 	<ul style="list-style-type: none"> • Bug in C after ASCII conversion • Eg. mum (E0041164) • Eg. muTBS (E0659182)
test_dmv.log	Retrieve derivational variants	<ul style="list-style-type: none"> • lvg -f:Ln:d -kd:3 • lvg -f:L:d -kd:3 • toAscii 	<ul style="list-style-type: none"> • 419,502 to 732 • 732 are difference of fact (see E-mail) 	<ul style="list-style-type: none"> • TBD 	<ul style="list-style-type: none"> • Consider Inflections • case sensitive? • Fact (582 vs. 4559) • Min. term length • Min. stem length • Non-LEXICON term?
test_dmvR.log	TBV	<ul style="list-style-type: none"> • lvg -f:Ln:d -kd:3 • ToAscii • lvg -f:E 	<ul style="list-style-type: none"> • TBD 	<ul style="list-style-type: none"> • TBD 	<ul style="list-style-type: none"> • TBD
test_lfc.log	Retrieve category	<ul style="list-style-type: none"> • lvg -f:Ln 	<ul style="list-style-type: none"> • 26 to 26 • All 26 (non-Ascii) 	<ul style="list-style-type: none"> • 260 • All 260 (non-Ascii) 	<ul style="list-style-type: none"> • Ascii issue I: the ASCII form of non-ASCII citation is not one of the spelling variants
test_vcf.log	TBV	<ul style="list-style-type: none"> • lvg -f:Ct • lvg -f:i:s • toAscii 	<ul style="list-style-type: none"> • TBD 	<ul style="list-style-type: none"> • TBD 	<ul style="list-style-type: none"> • TBD
test_vf.log	Retrieve spelling variants of inflectional term of input term	<ul style="list-style-type: none"> • lvg -f:i:s • toAscii 	<ul style="list-style-type: none"> • 40248 to 126823 	<ul style="list-style-type: none"> • TBD 	<ul style="list-style-type: none"> • Case: fact • Case: uncount noun (276, 014)
Test_glR.log	Retrieve	<ul style="list-style-type: none"> • lexAccess 	<ul style="list-style-type: none"> • 8,277 to 	<ul style="list-style-type: none"> • TBD 	<ul style="list-style-type: none"> • TBD

	lexical records	?			
??	Approximate match	??	• TBD	• TBD	• TBD
??	Retrieve base form (citation form?) begin with xxx	<ul style="list-style-type: none"> • Lvg -f:Lp • ToAScii 	• TBD	• TBD	• TBD