

This document provides an overview of the Corpus and Repository of Writing (Crow)'s interface. Organizations planning a web-based corpus search interface can use this document to evaluate whether Crow's open-source software can be adapted or used as a model.

This document covers:

- 1. Design & usability of the interface
- A case study in adapting the software for a similar, multilingual corpus
- 3. Technical specifications of the software

Design & Usability: Main navigational features

- filter by draft, assignment, and country (e.g., find first drafts of synthesis papers written by Indian students)
- search by exact word or word family (e.g., "argue, argues, argued, arguing")
- See highlighted writing excerpts

Interface Design

- **Dataset:** a large-scale learner corpus of English writing samples from university foundational writing courses, as well as pedagogical materials used in those courses
- User audience: corpus researchers, writing teachers, and students
- Access: registration required; different tiers of access

argue state	Search	Reset Export	D		
✓ Advanced Search					
Tip: To search by phrase, place multiple words in quotation marks					
Word form handling: • Exact word	○ Word family (e.g., leave/leaves/leav	vina/left)			
•	, ,				
Pritter by ID					
Matching texts: 573 Word c	count in search range: 2.065.476	Embed	🗹 Display metadata		
Search	h result counts, ra		ed		
			Texts containing term		
			-		
argue	306	148.15	226		
state	675	326.80	413		
(-0-				
(combined)	981	474.95	573		
Chausing matching tests (op of 570	_				
Showing matching texts 1-20 01 573	Exce	rpted sear	ch results		
has a surface and the sheet BY a second	tenden in en en en la deux <mark>en en e</mark> black (b. te		and the second		
	icular, some scholars <mark>argue</mark> that it is	unnecessary for non-r	native speakers to		
sound like native speakers of Er	,				
	Advanced Search Tip: To search by phrase, place mult Word form handling: Exact word Multiple word handling: Contains Filter by ID. Matching texts: 573 Word c Search query Instanc argue	Advanced Search Tip: To search by phrase, place multiple words in quotation marks Word form handling: Exact word Word family (e.g., leave/leaves/lea	Advanced Search Tip: To search by phrase, place multiple words in quotation marks Word form handling: Exact word Word family (e.g., leave/leaves/leaving/left) Multiple word handling: Contains any word Contains all words Filter by ID. Matching texts: 573 Word count in search range: 2,065,476 Embad Search result counts, raw & norm Search query Instances in matching texts Normed (per 1 million) argue 306 148.15 state 675 326.80 (combined) 981 474.95		

Figure 2: Individual page view of a pedagogical material

Syllabus: ENGL 106INTL	Download Original		
PDF Plaintext	106i_NA_SL_969_PRD (256.85 KB)		
2	 Instructor materials related to this resource Related course materials 		
FIRST-YEAR COMPOSITION FOR INTERNATIONAL STUDENTS Course Information Course Number:	Document type Syllabus		
CRN: Course Meeting Time: Classroom: Instructor's Nnme: Instructor's Office: Instructor's Office:	Institution Purdue University Term Spring 2018		
Instructive's Office Hours: Instructive's Office Hours: ENG106 is the standard 4-credit hour composition course for students at Purdue. The course provides students with the opportunity to interpret and compose in both digital and print media across a variety of forms. Students enging in active learning, which indudes class discussion, learning in small groups, problem solving, and peer review. ENG 106 is grounded in the idea that writing provides an outlet for sharing and developing ideas, Ecilitatics understanding across different conventions, genres, groups,	ENGL 106INTL ENGL :106INTL :is the standard 4-credit hour composition course for international students at		

Figure 1: Screenshot of the main corpus search interface



Case study: Adapting the software for the MACAWS corpus

MACAWS is a corpus and repository for World Languages. Though their datasets differ – most notably, MACAWS includes Russian and Portuguese language texts – Crow's search interface fit the needs of the MACAWS corpus.

Institution	to write	Search Reset	► Target Language	escrever Search Reset Export ①	
· Year	Advanced Search		► Course	Advanced Search	
Semester	Matching texts: 10,851 Word count in search ran	ge: 10,155,120 Embed	► Macro Genre	Matching texts: 56 Word count of subcorpus: 287,379 ① Embed	
Course	Search query Instances in matching texts ①	Normed (per 1 million) ① Texts containing term	► Topic	Search query Instances in matching texts ① Normed (per 1 million) ① Texts containing term	
Assignment	to 310,329	30.558.87 10.851	► Draft	escrever 74 257.50 56	
Draft	write 10.927	1.076.01 2,780	Mode of Assignment	Showing matching texts 1-20 of 56	
College	(combined) 321,256	31,634.88 10,851	►La		
Country	Showing matching texts 1-20 of 10.851			tte de sua amiga. " Eu decid <mark>escrever</mark> sua dedicatória porque é parte da história. Tristemente eu encontrei as fotografias de minha avó despois de que ela morreu, mas minha mãe e meu pai me contaram alguma coisa sobre a fotografia que eu encontrei muito interessante. Minha avó e	
Gender	hat do non-native speakers say about who they want to sound like?", sung, 2013) There are appear many kinds of English. But most people believe that t			Term: Spring 2018 Course: PORT 325 Assignment: Media (Writing) Draft: Final Wordcount: 335	
Program	Term: Spring 2018 Course: ENGL 106 ① Assignment: Position Argument ① Draft: 2 Gender: F TOEFL: 73			uriosidades que me cativaram. Havia uma maquina de escrever antiga e abandonada fora de uma casa com flores crescendo nos espacos entre a chaves e a rua. O que gostei de a rua porque verdadeiramente é	
Year in School				com nores crescendo nos espaços entre a chaves e a rua. O que gostel de a rua porque verdadeiramente e universal. Em todas partes há ruas e eles ajudam a pessoas a encontrar onde estão. Para mim, eles t	
TOEFL Score	odel in English courses If you are given a choice to common chef as your teacher, I am pretty sure y ls first which means when they are learning how to			Term: Spring 2018 Course: PORT 325 Assignment: Media (Writing) Draft: Final Wordcount: 263	
	precisely. I confidently support the points b			bom, que você quase podia sentir o gosto na boca. Escrever essa memória está me deixando com água	
	Term: Fall 2017 Course: ENGL 106 (i) Assignment: Por	sition Argument (i) Draft: 2 Gender: F TOEFL: 87		na boca. Pedimos uma grande mesa no meio do restaurante para que pudéssemos ver a vista da praia. Eu lembro de olhar para o menu com tanta fome que eu guería cada prato que eles serviram. O prato que	

In terms of data and metadata, they differ in:

Corpus & Repository of Writing https://crow.corporaproject.org/	Multilingual Academic Corpus of Assignments: Writing and Speech https://macaws.corporaproject.org
Focused on Second Language Writing, particularly English as Second Language.	Focused on Second Language Writing and Speech, particularly World Languages such as Portuguese,and Russian so far.
Repository availability: Yes! Tabs: for instructors, students, researchers, and developers available.	Repository availability: Soon! Tabs: not available.
Access: Sign in required Request access <u>here</u> .	Access: Sign in required Request access <u>here</u> .
Faceted metadata search Institution Year Semester Course Assignment: <i>Categorized by writing genre</i> Draft College Country Gender Program Year in School TOEFL Scores On Crow you can do: Keyword/phrase search Viewing individual texts Embedding search results (iDDL) Exporting results	Faceted metadata search Target language Course Macro genre Topic Draft Mode of the assignment: <i>Speaking or Writing</i> L1 On MACAWS you can do: Keyword/phrase search Viewing individual texts Embedding search results (iDDL) Exporting results