ALVEO The Human Communication Science Virtual Laboratory
building on HCSNet (an ARC research network)

Presented by: Peter Sefton

Steve Cassidy** Dominique Estival* Peter Sefton*, Jared Berghold*** Denis Burnham*

*University of Western Sydney, Australia;
** Macquarie University, Australia;
***Intersect Australia
Funding

Alveo acknowledges funding from the NeCTAR project
http://www.nectar.org.au NeCTAR is an Australian Government project
conducted as part of the Super Science initiative and financed by the
Education Investment Fund.
Above and Beyond Speech, Language and Music

A Virtual Lab for Human Communication Science

Get Started with Alveo

The Alveo provides on-line infrastructure for accessing human communication data sets (speech, texts, music, video, etc.) and for using specialised tools for searching, analysing and annotating that data.

🌟 Data Discovery Interface

Browse and search collections, view documents and create lists of items for further analysis. The Data Discovery Interface provides the jumping-off point for further analysis using the Galaxy Workflow Engine, the NeCTAR Research Cloud, the R statistical package or any other preferred tool or platform. A fully featured API underpins the Data Discovery Interface, providing opportunities to extend the functionality of the Virtual Laboratory.

Go to Alveo web app »

➡️ Galaxy Workflow Engine

Initially targeted at genomics researchers, Galaxy is a scientific workflow system which is largely domain agnostic. The Galaxy Workflow Engine provides Alveo users with a user-friendly interface to run a range of text, audio and video analysis tools. Workflows defining a sequence of steps in an analysis can be created and then shared with other researchers.

Go to Galaxy »

🌐 Accessible

Accessible to non-technical researchers via workflow tools, stored protocols, and interactive GUIs, while also allowing non-experts to interact with specialist tools.

🌐 Interoperable

Interfaces are provided to the UIMA Java framework, Python and NLTK and the Emu/R environment. Annotations are stored using RDF following a model.

🌐 Sustainable

13 universities, 3 organisations, and 47 key investigators have provided support for sustained operational development and further capability development.
## Contributing Partners

<table>
<thead>
<tr>
<th>Gold</th>
<th>Silver</th>
<th>Bronze</th>
<th>Contributor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flinders</td>
<td>ANU</td>
<td>Griffith</td>
<td>ASSTA</td>
</tr>
<tr>
<td>Macquarie</td>
<td>Intersect</td>
<td>La Trobe</td>
<td>UTAS</td>
</tr>
<tr>
<td>RMIT</td>
<td>UC</td>
<td>NICTA</td>
<td></td>
</tr>
<tr>
<td>UNE</td>
<td>UMELB</td>
<td>UNSW</td>
<td></td>
</tr>
<tr>
<td>UWS</td>
<td>USYD</td>
<td>UWA</td>
<td></td>
</tr>
</tbody>
</table>
Development partner: Intersect

Intersect Development Team:

Ilya Anisimoff
Jared Berghold
David Clarke
Georgina Edwards
Karen El-Azzi
Gabriel Gasser
Matthew Hillman
Chris Kenward
Nasreen Sharique
Kali Waterford
Elyse Wise
Marc Ziani de Ferranti
Shuqian Hon
Sean Lin
Theeban Soundararajan
Vincent Tran
Stanley Hon
Pierre Estefahan
Simon Yin
Watch the video!

Colloquialism By Arthur Delbridge

There seem to me to be two questions about colloquialism for this Council: first, what is it? Are there some characteristics of a word or a phrase that would let you say confidently this is a colloquialism, or is not a colloquialism? And supposing the answer to the first is "yes" - here's the second question: in what circumstances is the use of colloquialism either good or bad? What place does colloquialism hold in communicative competence?

So let me start with the question of definition. And I shall work my way into this question first entirely as a lexicographer, concerned only with words and phrases as they may be listed and labelled in a dictionary.

As you know, dictionaries often use colloquial as a status label, along with other labels like formal, slang, academic, educated, informal, vulgar.
## Data Discovery

### Licence Agreements

#### Review and Acceptance of Licence Terms

Manage your subscriptions to collections.

<table>
<thead>
<tr>
<th>Collection or Collection List</th>
<th>Collections</th>
<th>Owner</th>
<th>State</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>AusNC</td>
<td>8</td>
<td><a href="mailto:jared@intersect.org.au">jared@intersect.org.au</a></td>
<td>Not Accepted</td>
<td>Preview &amp; Accept Licence Terms</td>
</tr>
<tr>
<td>PARADISEC</td>
<td>2</td>
<td><a href="mailto:jared@intersect.org.au">jared@intersect.org.au</a></td>
<td>Not Accepted</td>
<td>Preview &amp; Accept Licence Terms</td>
</tr>
<tr>
<td>avoizes</td>
<td>1</td>
<td><a href="mailto:jared@intersect.org.au">jared@intersect.org.au</a></td>
<td>Accepted</td>
<td>Review Licence Terms</td>
</tr>
<tr>
<td>austalk</td>
<td>1</td>
<td><a href="mailto:jared@intersect.org.au">jared@intersect.org.au</a></td>
<td>Not Accepted</td>
<td>Preview &amp; Accept Licence Terms</td>
</tr>
</tbody>
</table>
Discover data via metadata facets
Major functions of the repository component

Data Repository
( Hydra/Fedora + Sesame Triple Store )

Equivalent

API ➔ Hydra Web discovery service

Discovery

Access

Item List Manager ➔ Stable refs
Sesame RDF search ➔ Indexes
Solr Text/Field search ➔ Indexes
License/Access manager ➔ Controls access
Fedora Repository ➔ Controls access
Q. What do you call something that automatically prepares data for ingest?
• A: Robochef
Sivusto Rakenteilla

UNDER CONSTRUCTION
Architecture: Discovery leads to data

### Documents

<table>
<thead>
<tr>
<th>Filename</th>
<th>Type</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-200.txt</td>
<td>Original</td>
<td>6.2 kB</td>
</tr>
<tr>
<td>3-200-raw.txt</td>
<td>Raw</td>
<td>6.2 kB</td>
</tr>
<tr>
<td>3-200-plain.txt</td>
<td>Text</td>
<td>6.0 kB</td>
</tr>
</tbody>
</table>
Including various media

mitcheldelbridge:S1232s1

Primary Document

Item Details

Identifier: S1232s1
Collection: mitcheldelbridge
Mode: unspecified
Speech Style: unspecified
Interactivity: unspecified
Communication Context: unspecified
Discourse Type: interactive_discourse
Post discovery: compile your own stable “Item Lists”

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Title</th>
<th>Created Date</th>
<th>Type(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cocee:2-011</td>
<td></td>
<td>1826</td>
<td>Original, Raw, Text</td>
</tr>
<tr>
<td>cocee:4-288</td>
<td></td>
<td>1893</td>
<td>Original, Raw, Text</td>
</tr>
<tr>
<td>cocee:2-271</td>
<td></td>
<td>1843</td>
<td>Original, Raw, Text</td>
</tr>
<tr>
<td>cocee:3-022</td>
<td></td>
<td>1851</td>
<td>Original, Raw, Text</td>
</tr>
<tr>
<td>cocee:2-330</td>
<td></td>
<td>1848</td>
<td>Original, Raw, Text</td>
</tr>
<tr>
<td>cocee:2-259</td>
<td></td>
<td>1842</td>
<td>Original, Raw, Text</td>
</tr>
<tr>
<td>cocee:4-228</td>
<td></td>
<td>1890</td>
<td>Original, Raw, Text</td>
</tr>
<tr>
<td>cocee:1-103</td>
<td></td>
<td>1806</td>
<td>Original, Raw, Text</td>
</tr>
<tr>
<td>cocee:3-200</td>
<td></td>
<td>1860</td>
<td>Original, Raw, Text</td>
</tr>
<tr>
<td>cocee:2-280</td>
<td></td>
<td>1844</td>
<td>Original, Raw, Text</td>
</tr>
<tr>
<td>cocee:3-282</td>
<td></td>
<td>1872</td>
<td>Original, Raw, Text</td>
</tr>
</tbody>
</table>
Concordance: a tool run on an Item List

cooe:3-022  Found 1 match

more years to run out, it will then according to the general rate of letting

cooe:1-249  Found 5 matches

colony, and if they did not apply then he would beg to enquire when did

as the Common Law Officer. He was then told the Acts of Parliament prescribed a

were not in force in this colony, then the application of the former occasion was

penalty for his neglect. It was clear then that there was a legal right; it

without doubt apply here. Upon general principles then the court of sessions having power to
API Access: Get a key

No API Key generated

Generate API Key

Copy to Clipboard
Download API Key

Saved Searches
History
My Account
My Licence Agreements
API Key
Report An Issue
Logout
Copy-paste access to data

Copy the following code into your R environment and download the API token file. Make sure you have all the required R packages installed prior to execution.

Save the following file to your home directory
Linux or Unix: /home/<user>
Mac: /Users/<user>
Windows: C:\Users\<user>

Download API key config file

```r
library(emuSX)
item_list = readItemList('http://ic2-hcsvla-bl-staging1-vm.intersect.org.au/item_lists/129.json')
```
Workflow: Galaxy

The following job has been successfully added to the queue:

1: HCSVLAB data

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.
Chain processes on Item Lists - eg

[Tokenizer] -> [Frequency List]
source1

agroup1

dada:annotates

anno1

dada:partof

type: Annotation

ptb: PartOfSpeech

NNP

ptb:POS

region1

dada:targets

dada:start

4

dada:end

7
Reuse

Get ahead on your repository.

Multi-Purpose Repository Solutions
Flexible User Interfaces
Durable Digital Asset Management

powered by Galaxy
Reproducible Research
"I really liked using the system and the instructions were very easy to use and the system easy to navigate. [...] This platform would be very useful for my research."

--Tester