

Text Analytics: Investigating duplication in units of competency

Dr Alexander Ferry^{*}, Mr Paul Edwards^{*},
Dr Simon Stratton[^] and Mr Michael Wyndham[^]

Australian Government Department of Education and
Training

VET Data Analytics^{*} and Industry Policy[^]

Overview of presentation

1. Current policy objectives.
2. Development of a VET unit search engine.
3. Example application 1: demonstrating machine learning by grouping units related to “environmental sustainability”.
4. Example application 2: grouping units related to “team work and communication”, a detailed validation.
5. Conclusions.

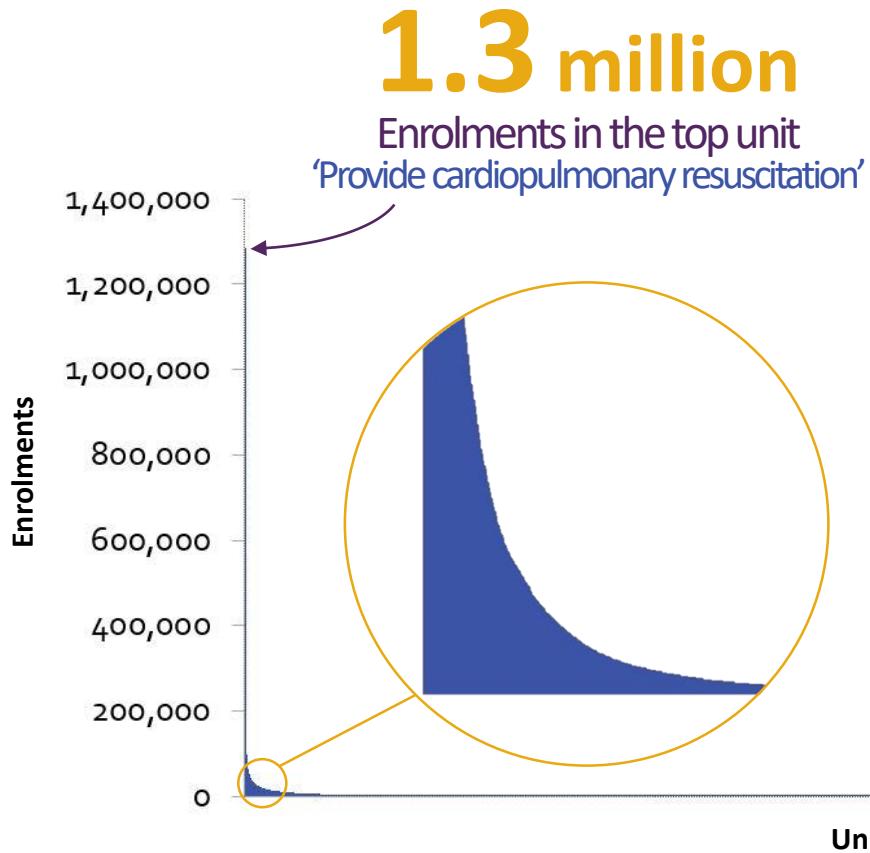
Current Policy Objectives

COAG Industry and Skills Council ministers have agreed to reforms that:

- remove obsolete and superfluous training products,
- create units of competency that are owned and used by multiple sectors.

Given the size and complexity of the national register, the department is supporting the Australian Industry and Skills Committee to progress these reforms.

Current Policy Objectives



A **unit of competency** (unit) refers to a specific standard of performance required in the workplace. They can loosely be thought of as *subjects*. This chart presents the distribution of 2017 unit enrolments for all units that were current at some point in 2017.

12,647

Number of units with
at least 1 enrolment

17,451

Total number of units

23 million

Enrolments in all units

5%

Enrolments
in top unit

50%

Enrolments in
top 232 units

90%

Enrolments in
top 1,974 units

Development of a VET unit search engine

The VET Data Analytics and Industry Policy teams developed a tool to find duplicate units of competency.

The tool can be broken into two parts:

- Search Engine: Find relevant unit of competency
- Clustering Algorithm: Group similar units of competency



[Home](#) > [Search](#) > Training search

Nationally recognised training search

Your search for **Communicate in the workplace.** has returned **7** results. If you are unable to find an appropriate result try restricting your search criteria.

[Modify Search](#)

[New Search](#)

▼ Units of competency (7 results)

Code	Title ▲	Usage recommendation
BSBCMM201	Communicate in the workplace	Current
AMPCOR205	Communicate in the workplace	Current
MSMSUP102	Communicate in the workplace	Current
ICPSUP262	Communicate in the workplace	Current
PUACOM001C	Communicate in the workplace	Current
RIICOM201D	Communicate in the workplace	Current
SIRXCOM001	Communicate in the workplace to support team and customer outcomes	Current



Items per page **10** | 20 | 50 | 100

Displaying items 1 - 7 of 7

[Back to top](#)




[Home](#) [Training](#) [BSBCMM201](#)


Unit of competency details

BSBCMM201 - Communicate in the workplace (Release 1)

Summary

Releases:	Release 1 (this release)	Status: Current	Release date: 25/Mar/2015
-----------	--------------------------	------------------------	----------------------------------

Usage recommendation: **Current**

Mapping:	Notes	Date
Supersedes and is equivalent to BSBCMM201A - Communicate in the workplace	Updated to meet Standards for Training Packages	24/Mar/2015

Companion volumes: Unit Of competency

- [Companion Volume implementation guides are found in VETNet](#)
- Assessment requirements
- [Companion Volume implementation guides are found in VETNet](#)

Delivery: [Find RTOs approved to deliver this unit of competency](#).

Training packages that include this unit

Code	Title	Release
SIT	Tourism, Travel and Hospitality Training Package	1.0-1.1
PRM	Pulp & Paper Manufacturing Industry Training Package	1.0
NWP	National Water Training Package	1.0
MAR	Maritime Training Package	3.0-4.0
HLT	Health	2.0-3.1
DIF	Defence Training Package	2.0
CIA	Creative Arts and Culture Training Package	2.0
CPD07	Property Services Training Package	14.4-14.5
CSC	Community Services	2.0-3.0
BSB	Business Services Training Package	1.0-1.2,2.0

Qualifications that include this unit

Code	Title	Release
SIT31116	Certificate III in Asian Cookery	1
SIT31016	Certificate III in Patisserie	1
SIT30916	Certificate III in Catering Operations	1
SIT30816	Certificate III in Commercial Cookery	1
SIT30616	Certificate III in Hospitality	1-2
SIT20516	Certificate II in Asian Cookery	1
SIT20416	Certificate II in Kitchen Operations	1
SIT20316	Certificate II in Hospitality	1-2
SIT20216	Certificate II in Holiday Parks and Resorts	1-2
SIT20116	Certificate II in Tourism	1-2

1 2 3 4

Items per page 10 | 20 | 50 | 100

Displaying items 1 - 10 of 31

Skill sets that include this unit

Code	Title	Release
BSBSS00041	Innovation Practice Skill Set	1

Classifications

Scheme	Code	Classification value
ASCED Module/Unit of Competency Field of Education Identifier	120505	Work Practices Programmes

Content

Compare: [Compare content of this unit of competency with other releases or training components](#)

Download: [Unit Of competency \(1.26 MB\)](#) [Unit Of competency \(165.43 KB\)](#)

[Assessment requirements \(1.26 MB\)](#) [Assessment requirements \(147.34 KB\)](#)

Unit Of competency

Modification History

Release	Comments
Release 1	This version first released with BSB Business Services Training Package version 1.0

Application

This unit describes the skills and knowledge required to communicate in the workplace including gathering, conveying and receiving information and completing routine written correspondence.

It applies to individuals who perform a range of routine workplace communication tasks using a limited range of practical skills and fundamental knowledge of effective listening, questioning and non-verbal communication in a defined context under direct supervision or with limited individual responsibility.

No licensing, legislative or certification requirements apply to this unit at the time of publication.

Unit Sector

Communication – Interpersonal Communication

Elements and Performance Criteria

ELEMENT	PERFORMANCE CRITERIA
<i>Elements describe the essential outcomes.</i>	<i>Performance criteria describe the performance needed to demonstrate achievement of the element.</i>
1 Gather, convey and receive information and ideas	1.1 Collect information to achieve work responsibilities from appropriate sources 1.2 Use method/s and/or equipment to communicate appropriate ideas and information to the audience 1.3 Use effective listening and speaking skills in verbal communication 1.4 Seek input from internal and external sources to develop and refine new ideas and approaches 1.5 Respond to instructions or enquiries promptly and in accordance with organisational requirements
2 Complete workplace documentation and correspondence	2.1 Present written information and ideas in clear and concise language to ensure the intended meaning of correspondence is understood by the recipient 2.2 Draft and present correspondence within designated time lines 2.3 Ensure presentation of written information meets organisational standards of style, format and accuracy 2.4 Complete workplace forms and documentation in a clear, concise and easy to read format
3 Communicate in a way that responds positively to individual differences	3.1 Value all individuals and treat them with respect, courtesy and sensitivity 3.2 Take into consideration cultural differences in all verbal and non-verbal communication 3.3 Use communication to develop and maintain positive relationships, mutual trust and confidence 3.4 Make efforts to use basic strategies to overcome language barriers 3.5 Ensure that behaviour is consistent with legislative requirements, enterprise guidelines and/or social protocols

Foundation Skills

This section describes language, literacy, numeracy and employment skills incorporated in the performance criteria that are required for competent performance.

Skill	Performance Criteria	Description

Development of a VET unit search engine

Pre-processing

This unit describes the skills and knowledge required to communicate in the workplace including gathering, conveying and receiving information and completing routine written correspondence.

It applies to individuals who perform a range of routine workplace communication tasks using a limited range of practical skills and fundamental knowledge of effective listening, questioning and non-verbal communication in a defined context under direct supervision or with limited individual responsibility.

No licensing, legislative or certification requirements apply to this unit at the time of publication.

Development of a VET unit search engine

Pre-processing

communicate in the workplace including gathering conveying and receiving information and completing routine written correspondence it applies to individuals who perform a range of routine workplace communication tasks using a limited range of practical skills and fundamental knowledge of effective listening questioning and non verbal communication in a defined context under direct supervision or with limited individual responsibility

Development of a VET unit search engine

Pre-processing

commun+ workplac+ include+ gather convey receiv+ inform
complet+ written correspond appli+ individu+ perform rang+
workplac+ commun+ task use rang+ skill knowledg+ listen
question commun+ context supervis+ respons+

{communicate, communication} → commun+

Development of a VET unit search engine

Weighting

$$TF = \frac{\text{frequency}}{\text{maximum frequency}}$$

commun+ workplac+ include+ gather convey receiv+ inform
complet+ written correspond appli+ individu+ perform rang+
workplac+ commun+ task use rang+ skill knowledg+ listen
question commun+ context supervis+ respons+

Term	Count	TF
“commun+”	3	1.00
“workplac+”	2	0.67
“rang+”	2	0.67
Others	1	0.33

Development of a VET unit search engine

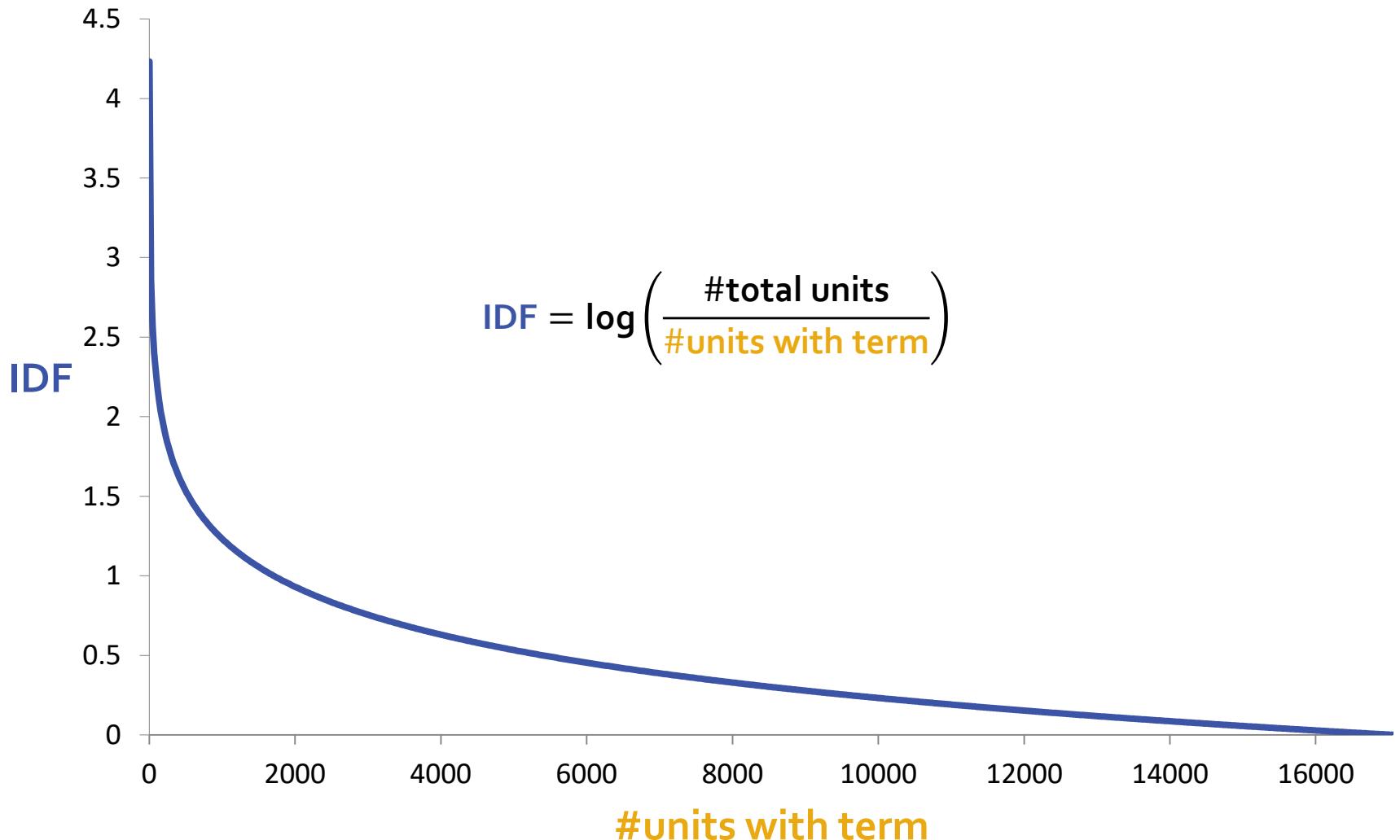
Weighting

$$\text{IDF} = \log \left(\frac{\#\text{units}}{\#\text{units with term}} \right)$$

Term	Count	IDF
correspond	64	5.59
listen	286	4.09
convey	428	3.69
question	610	3.33
written	913	2.93
gather	941	2.9
receiv+	1228	2.63
context	3788	1.51
task	4252	1.39
supervis+	5099	1.21
Individu+	5677	1.1
complet+	5698	1.1
respons+	5837	1.07
workplac+	6315	0.99
commun+	6399	0.98
rang+	6652	0.94
knowledg+	7189	0.86
skill	7929	0.77
inform	8450	0.7
Include	8991	0.64
perform	9399	0.6
use	13306	0.25
appli+	15986	0.07

Development of a VET unit search engine

Weighting



Development of a VET unit search engine

Weighting

$$\text{TF-IDF} = \text{TF} \times \text{IDF}$$

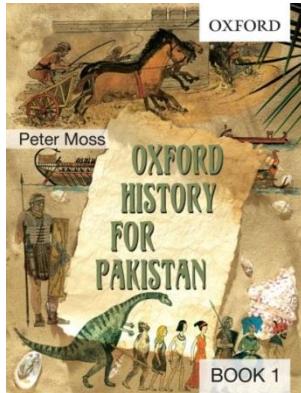
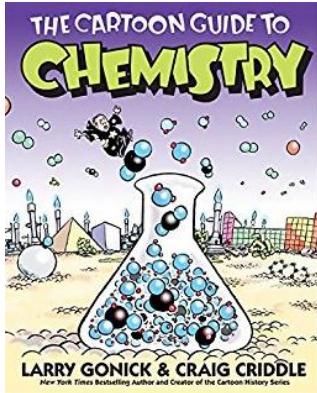
Term	TF	IDF	TF-IDF
correspond	0.33	5.59	1.84
listen	0.33	4.09	1.35
convey	0.33	3.69	1.22
question	0.33	3.33	1.10
commun+	1.00	0.98	0.98
written	0.33	2.93	0.97
gather	0.33	2.90	0.96
receiv+	0.33	2.63	0.87
workplac+	0.67	0.99	0.66
rang+	0.67	0.94	0.63
context	0.33	1.51	0.50
task	0.33	1.39	0.46
supervis+	0.33	1.21	0.4
individu+	0.33	1.10	0.36
complet+	0.33	1.10	0.36
respons+	0.33	1.07	0.35
knowledg+	0.33	0.86	0.28
skill	0.33	0.77	0.25
inform	0.33	0.70	0.23
Include	0.33	0.64	0.21
perform	0.33	0.6	0.20
use	0.33	0.25	0.08
appli+	0.33	0.07	0.02

Development of a VET unit search engine

Video

Development of a VET unit search engine

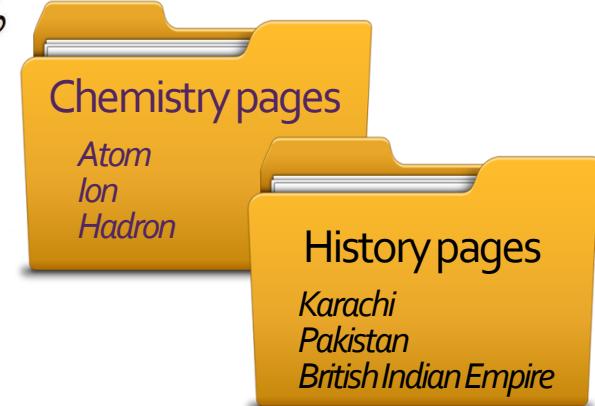
Machine Learning



Unstructured
text



Machine learning
technique



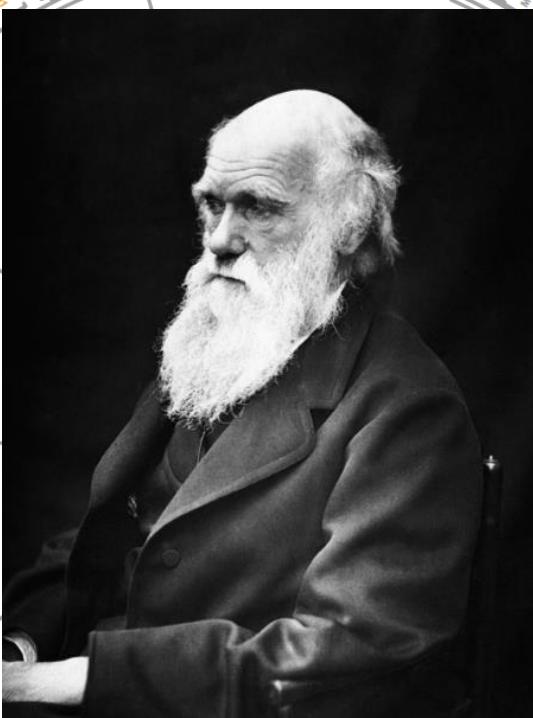
Example application 1: demonstrating machine learning by grouping units related to “environmental sustainability”

Tested our search tool and machine learning against the findings from consultants focussing on units related to “Skills for Environmental Sustainability”.

Found 573 relevant units, which included all but 8 of 213 units identified by the consultant.

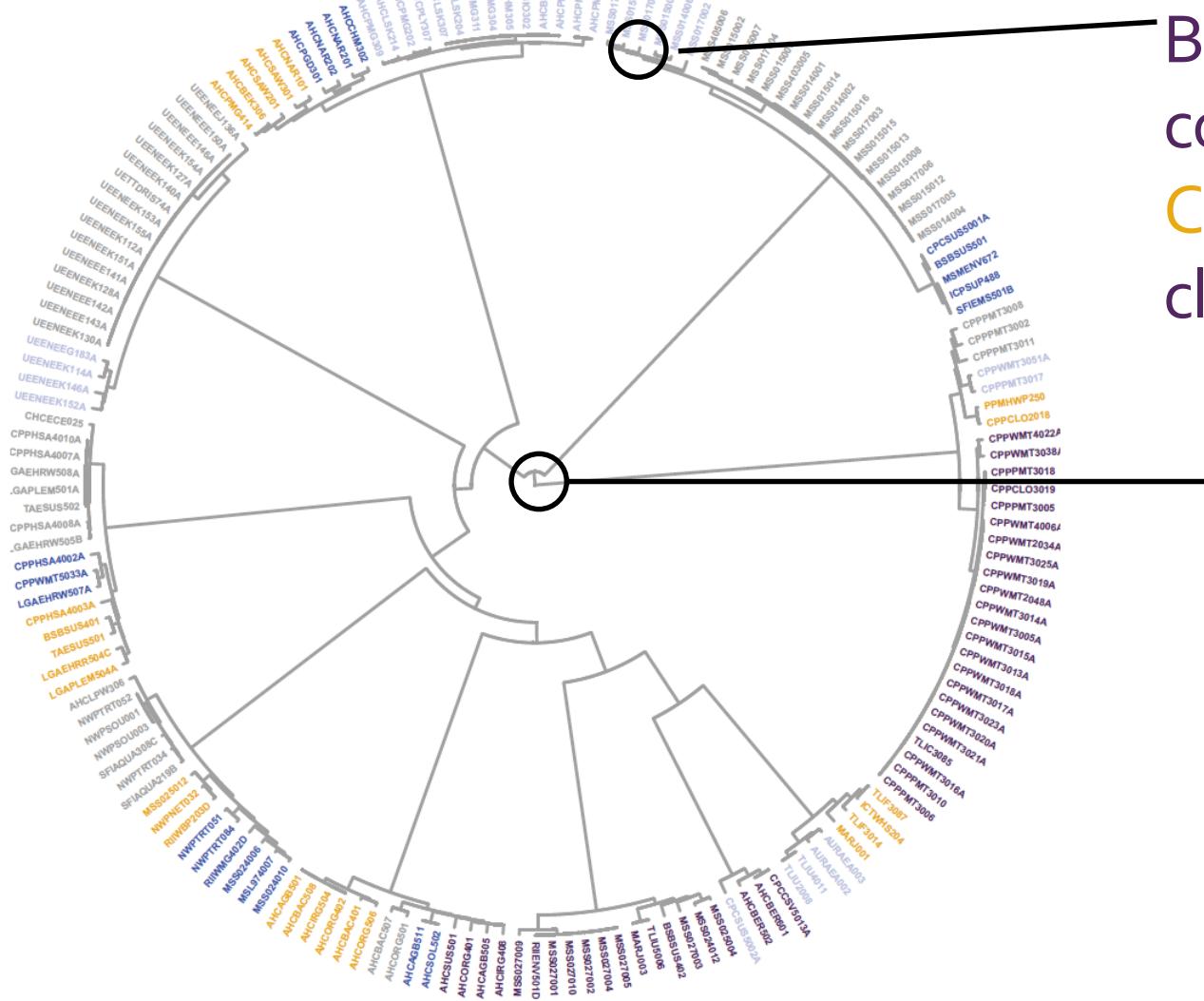
We used machine learning to help reveal how these 573 units were related.

Example application 1: demonstrating machine learning by grouping units related to “environmental sustainability”



This is a Dendrogram.
It's a bit like an
evolutionary tree,
illustrating how
similar units are
according to the LDA
algorithm.

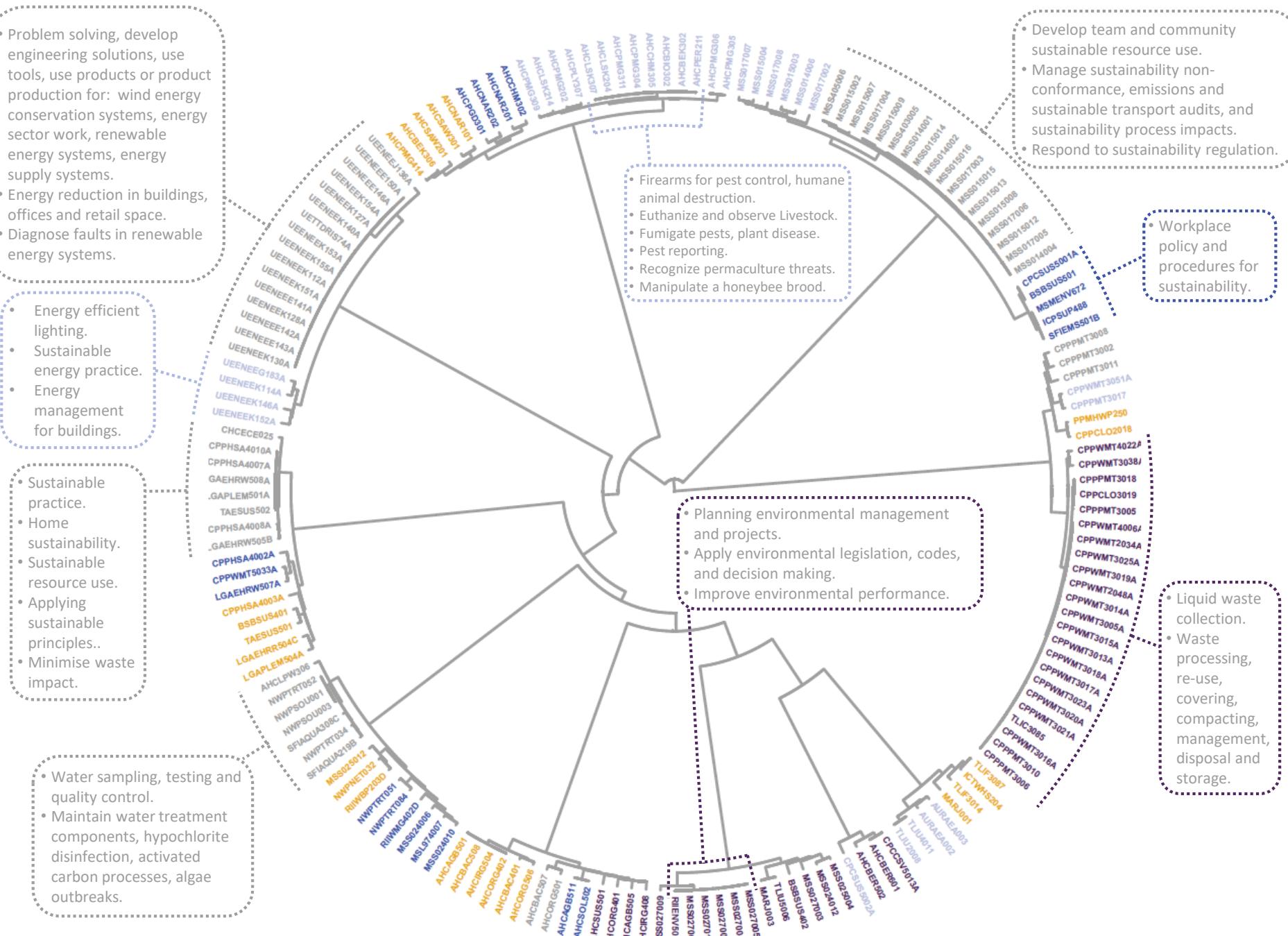
Example application 1: demonstrating machine learning by grouping units related to “environmental sustainability”



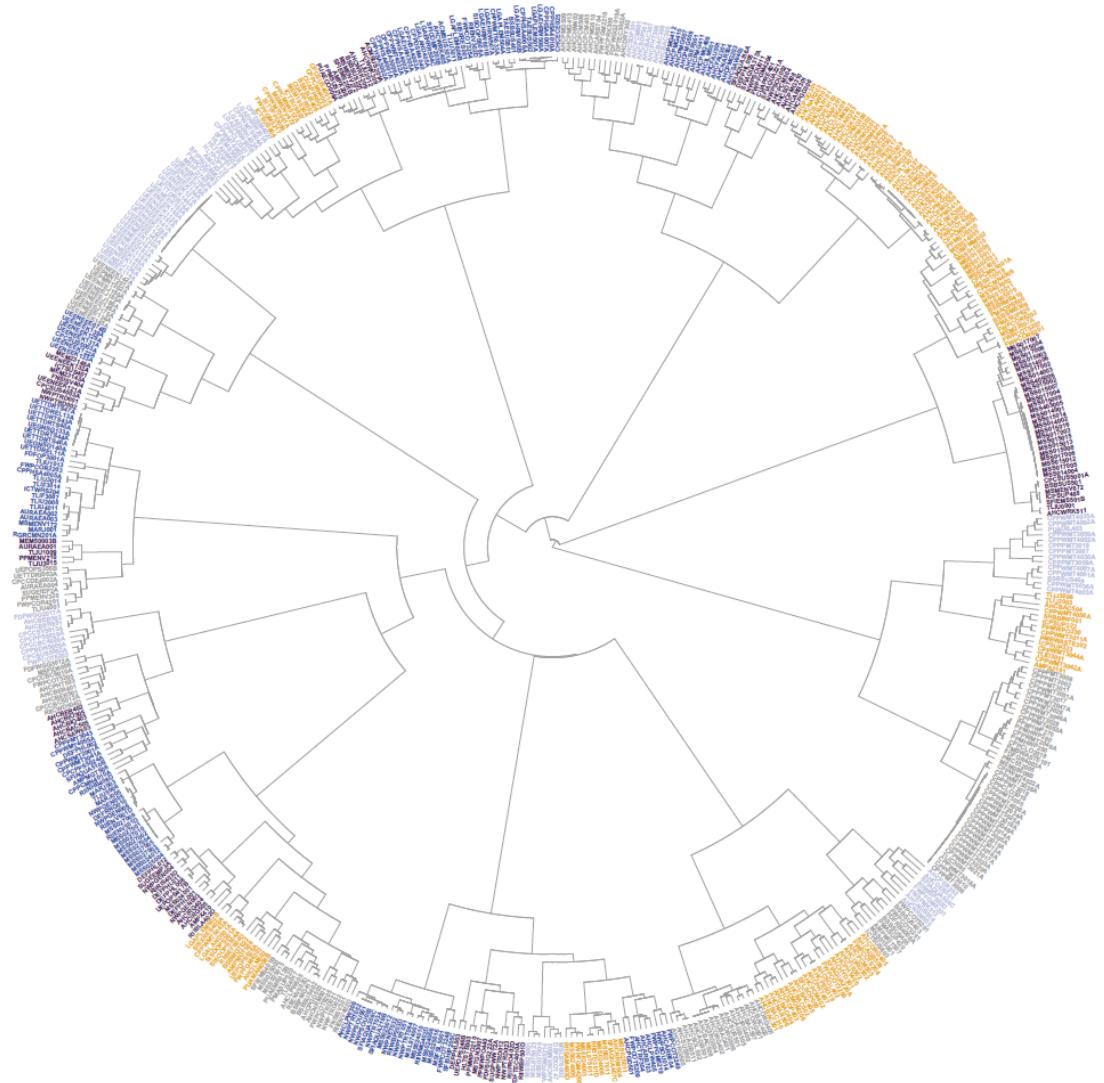
But here we'd only consider Budgies and Cockatoos to be closely related.

Here we'd consider Ants and Whales to be closely related and therefore grouped together.

Latent Dirichlet Allocation groups related units in “Environment and Sustainability” using unit text descriptions



Example application 1: demonstrating machine learning by grouping units related to “environmental sustainability”



A grouping for all 573 units related to “Skills for Environmental Sustainability”.

Illustrates the relationship between roughly 1100 pages of text.

Example application 2: grouping units related to “team work and communication”, a detailed validation

A second test case searched for units relevant to “team work and communication”, identifying 101 units.

Our Industry Advice branch read the descriptions for all 101 units, comparing the machine learning against expert human learning.

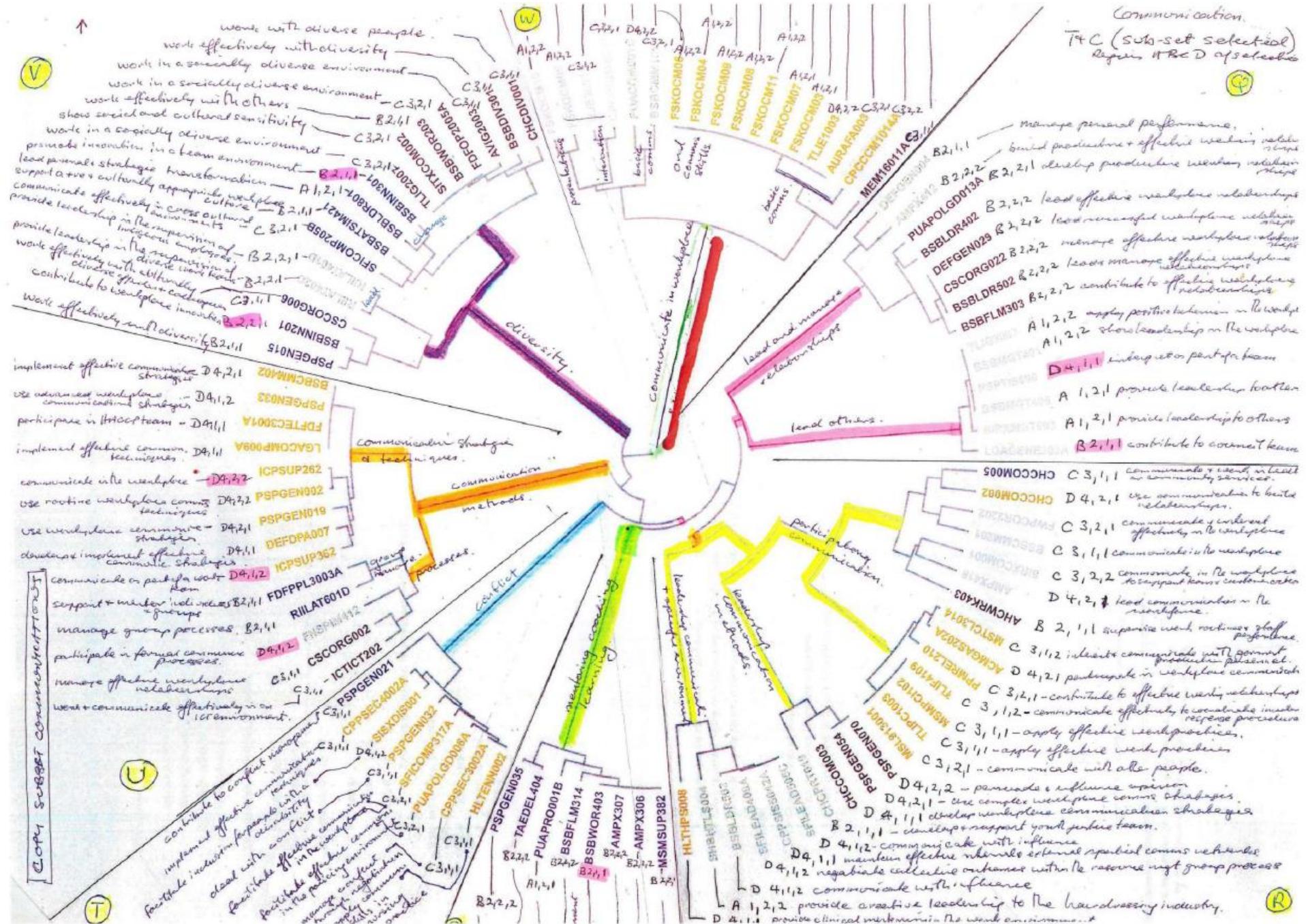
10 groups of units were clearly identified, covering the following general themes of:

- “diversity”, “lead and manage relationships”, “lead others”, “conflict”, “mentoring and coaching” and “leadership communication”.

Example application 2: grouping units related to “team work and communication”, a detailed validation

8 of these 10 groups were assessed to also have two meaningful subgroups.

Just 8 units were assessed as irrelevant to the search query (an estimated miss-classification rate of ~ 8%).



Conclusions

LDA has successfully identified units which are similar and revealed previously unknown groupings across units.

Allows automated comparison between proposed units and current units.

Can therefore identify where skills needs are shared between multiple industry sectors.

Conclusions

Could provide support for:

- identifying redundant and closely related units in skill areas shared by multiple industries.
- deleting redundant units and reducing the number of related units by establishing new units for cross industry use.

Conclusions

VET Data Analytics search algorithm and LDA grouping algorithm will soon be integrated with training.gov.au.

Available for use by all jurisdictions, Skilled Service Organisations and interested parties.