



CASCOT and the Coding of Occupations in European Surveys

1st DASISH Quantitative Workshop
Mannheim, 12-13 December 2012

Peter Elias and Margaret Birch
Institute for Employment Research
University of Warwick

Computer Assisted Structured Coding Tool

CASCOT



- Under this EC-funded project, IER has been contracted within work package 3 to develop a multi-lingual version of CASCOT to code job titles to ISCO 08.
- This is a large task, limited resources available so this will be a pilot project.
- The selected languages for the pilot are:
 - Dutch (Netherlands, Flemish-Belgium)
 - English
 - French (France, Walloon-Belgium, Switzerland)
 - German (Germany, Austria, Switzerland)
 - Slovak
 - Spanish

The work undertaken to date

Contact has been established with:

- Kea Tijdens (Universities of Amsterdam and Rotterdam)
- Manuel Munz (University of Bamberg Germany)
- Hartmut Minkel (Destatis Germany)
- Karol Václavík (Trexima Bratislava)
- David Hunter (ILO)
- Marc Houben (CBS Netherlands)
- Louis Meuric (INSEE France)
- Julio Nunez (INE Spain) – not yet responded
- Judith Pilliner (ONS UK)

A software development specification has been agreed with a development programmer

All the interface text has been extracted and placed into Excel for translation into the selected languages

Requirements for CASCOT development work

Three requirements for multi-language support for CASCOT:

- The user interface to be available in selected languages
- National language versions of ISCO08 structure
- Job titles in selected national languages indexed to ISCO08

Partnership arrangements for testing and development of language based coding rules

The user interface to be available in selected languages

ENGLISH	GERMAN	FRENCH	DUTCH	SPANISH	SLOVAK
File	Datei	fichier (d'archives)	Bestand		Súbor
Edit	Bearbeiten	éditer	Bewerken		Upraviť
Options	Optionen	options	Instellingen		Možnosti
Classification	Klassifikation	classification	Classificatie		Klasifikácia
Help	Hilfe	aide	Help		Pomoc
Input	Eingabe	input/entrée/saisie	Invoer		Vstup
No input file	keine Eingabedatei gefunden	aucun fichier saisie	Geen invoerbestand		Žiadny vstupný súbor
Record number	Nummer des Eintrags	Numéro de dossier (?)	Recordnummer		Číslo
Progress	Fortschritt	progrès	Voortgang		Priebeh
Text	Text	texte	Tekst		Text
Code	Code	code	On the 'code'-button: Codeer As column in the recommendation list: Code		Kód
Recommendations	Empfehlungen	recommandations	Aanbevelingen		Návrhy
Group title	Gruppenbezeichnung	titre du groupe	Code benaming		Názov skupiny
Best matching index entry	Genauester Indexeintrag	écriture index de la meilleur correspondence (?)	Best passende zoekterm		Najlepšie vyhovujúci záznam
Score	Score	score/cotation	Score		Skóre
Selection	Auswahl	choix/sélection	Selectie		Výber
Classification structure	Klassifikationsstruktur	structure de la classification	Classificatiestructuur		Štruktúra klasifikácie

Slovakian classification of occupations

(Štatistická klasifikácia zamestnaní SK ISCO-08 2,600 6-digit categories)

Extract:

2112	Meteorológovia
2112001	Letecký meteorológ
2112002	Meteorológ klimatológ
2112003	Hydrometeorológ
2112999	Meteorológ inde neuvedený
2113	Chemici (okrem chemického inžinierstva)
2113001	Chemik kvantovej chémie
2113002	Jadrový chemik
2113003	Analytický chemik
2113004	Fyzikálny chemik
2113005	Organický chemik
2113006	Anorganický chemik
2113999	Chemik inde neuvedený (okrem chemického inžinierstva)
2114	Geológovia a geofyzici
2114001	Geochemik
2114002	Pedológ, pôdoznalec
2114003	Mikropaleontológ
2114004	Paleontológ
2114005	Mineralóg, petrológ
2114006	Vedecký pracovník v geológii
2114007	Geológ (okrem banského geológa)
2114008	Hydrogeológ
2114009	Seizmológ
2114010	Geofyzik
2114011	Petrofyzik
2114999	Geológ a geofyzik inde neuvedený

French index to ISCO 08 (7,578 job titles,

INSEE)

Extract:

2111	2111	Aerodynamicist	Aérodynamicien
2111	2111	Astronomer	Astronome
2111	2111	Astronomer, radio	Astronome, radioastronome
2111	2111	Astrophysicist	Astrophysicien
2111	2111	Ballistician	Balisticien
2111	2111	Researcher, physics	Chercheur, physique
2111	2111	Hydrodynamicist	Hydrodynamicien
2111	2111	Physicist	Physicien
2111	2111	Physicist, clinical	Physicien, clinique
2111	2111	Physicist, nuclear	Physicien, physique nucléaire
2111	2111	Physicist, medical	Physicien, santé
2111	2111	Rheologist	Rhéologue
2111	2111	Thermodynamicist	Thermodynamicien
'2111	2111		Physicien, astronomie
'2111	2111		Physicien, électronique
'2111	2111		Thermicien
2112	2112	Climatologist	Climatologiste
2112	2112	Climatologist	Climatologue
2112	2112	Hydrometeorologist	Hydrométéorologiste
2112	2112	Forecaster, weather	Météorologiste
2112	2112	Meteorologist	Météorologiste
2112	2112	Meteorologist, climatology	Météorologue, climatologie
2112	2112	Meteorologist, weather forecasting	Météorologue, prévisions
'2112	2112		Ingénieur météorologiste, prévisions
'2112	2112		Météorologue

Correspondencia CN0-11 con CIUO08)

Extract	CNO-11		CIUO08	
	2411	Físicos y astrónomos	1	2111 Físicos y astrónomos
	2412	Meteorólogos	1	2112 Meteorólogos
	2413	Químicos	1	2113 Químicos
	2414	Geólogos y geofísicos	1	2114 Geólogos y geofísicos
	2415	Matemáticos y actuarios	1	2120 Matemáticos, actuarios y estadísticos
	2416	Estadísticos	1	2120 Matemáticos, actuarios y estadísticos
	2421	Biólogos, botánicos, zoólogos y afines	1	2131 Biólogos, botánicos, zoólogos y afines
	2422	Ingenieros agrónomos	1	2132 Agrónomos y afines
	2423	Ingenieros de montes	1	2132 Agrónomos y afines
	2424	Ingenieros técnicos agrícolas	1	2132 Agrónomos y afines
	2425	Ingenieros técnicos forestales y del medio natural	1	2132 Agrónomos y afines
	2427	Enólogos	1	2132 Agrónomos y afines
	2426	Profesionales de la protección ambiental	1	2133 Profesionales de la protección medioambiental
	2431	Ingenieros industriales y de producción	1	2141 Ingenieros industriales y de producción
	2461	Ingenieros técnicos industriales y de producción	1	2141 Ingenieros industriales y de producción
	2432	Ingenieros en construcción y obra civil	1	2142 Ingenieros civiles
	2462	Ingenieros técnicos de obras públicas	1	2142 Ingenieros civiles
	2437	Ingenieros ambientales	1	2143 Ingenieros medioambientales
	2433	Ingenieros mecánicos	1	2144 Ingenieros mecánicos
	2434	Ingenieros aeronáuticos	1	2144 Ingenieros mecánicos
	2463	Ingenieros técnicos mecánicos	1	2144 Ingenieros mecánicos
	2464	Ingenieros técnicos aeronáuticos	1	2144 Ingenieros mecánicos
	2435	Ingenieros químicos	1	2145 Ingenieros químicos
	2465	Ingenieros técnicos químicos	1	2145 Ingenieros químicos
	2436	Ingenieros de minas, metalúrgicos y afines	1	2146 Ingenieros de minas, metalúrgicos y afines
	2466	Ingenieros técnicos de minas, metalúrgicos y afines	1	2146 Ingenieros de minas, metalúrgicos y afines
	2439	Ingenieros no clasificados bajo otros epígrafes	1	2149 Ingenieros no clasificados bajo otros epígrafes

Dutch classification of occupations

(Wage Indicator Survey – 1,858 job titles: Kea Tidjens)

Extract:

2111010000000	Physicist	Fysicus, klinisch fysicus, gezondheidsfysicus
2112010000000	Meteorologist	Meteoroloog
2112020000000	Climatologist	Klimatoloog
2113010000000	Chemist	Chemicus
2113020000000	Petroleum chemist	Petrochemicus
2113030000000	Pharmaceutical chemist	Farmaceutisch chemicus
2113040000000	Textile chemist	Textielchemicus
2113050000000	Quality inspector chemical products	Kwaliteitscontroleur chemische producten
2113060000000	Quality inspector pharmaceutical products	Kwaliteitscontroleur farmaceutische producten
2114010000000	Geologist, geophysicist	Geoloog, geofysicus
2114040000000	Mining geologist	Mijnbouwgeoloog
2114050000000	Oil or gas geologist	Olie- of gasgeoloog
2114990000000	Physical scientists, all other	Natuurwetenschapper, anders

Dutch classification of occupations
(CBS – approx 5,500 job titles: Marc Houben)

Computer Assisted Structured Coding Tool

CASCOT Editor



Computer **A**ssisted **S**tructured **C**oding **T**ool

CASCOT Editor – how does it work?



CASCOT Editor – the tables of rules

Downgraded Words

A list of words which may be encountered in the text to be coded that are considered to be significantly less important than other words occurring in the classification structure or index.

Equivalent Word Ends

A list of word endings which may be encountered in the text to be coded and which are considered to be equivalent.

CASCOT Editor – the tables of rules

Abbreviations

A list of commonly encountered abbreviations for words and the meaning of those abbreviations. The abbreviations are expanded whenever they occur in the input text, the index or in other rules.

Replacement Words

This table defines a set of global word replacements. Word and phrases defined here are replaced whenever the words or phrase are encountered in input text, the index, or in other rules.

Conclusions

These rules define how Cascot should conclude when the defined word(s) or phrase is encountered in the input text. Ambiguous text is limited to a score of 39. When text marked as 'can not conclude' is encountered Cascot will not list any recommendations.

Default Rules

These rules define a set of words and phrases that should be scored as though they were a different word or phrase. To force Cascot to code a piece of text to a particular code a default rule can be added with the 'score as' words being one of the index entries from the classification.

Critical issues to be resolved

How will we test the software?

The software needs to be run under controlled conditions using job title information in the selected languages that has already been coded to ISCO08. This will generate rules to be incorporated into the classification files.

Is it too late for other countries to become involved?

The pilot will probably produce one or two versions of the software which work well. We will evaluate these to determine what worked well and what did not produce good results to learn lessons for future development. Where there is sufficient interest from another country we will encourage them to develop the software themselves on the basis of the lessons learnt.

Further information about CASCOT

[http://www2.warwick.ac.uk/fac/soc/ier/publications/
software/cascot](http://www2.warwick.ac.uk/fac/soc/ier/publications/software/cascot)