



| | |
|-------------------|--|
| Project no. | FP6-028987 |
| Project acronym | <i>EurOccupations</i> |
| Project title | Developing a detailed 7-country occupations database for comparative socio-economic research in the European Union |
| Instrument: | STREP |
| Thematic Priority | PRIORITY 7, Research Area 3.2.2. |

**Development of
associated software for
EurOccupations database
Third Reporting Period – D26**

| | |
|----------------------------|--|
| Project coordinator name: | Erasmus University Rotterdam |
| Start date of the project: | from 01.05.2006 to 30.04.09 (36 months) |
| Period covered: | from 01.05.08 to 30.04.09 |
| Report Preparation Date: | 30.04.2009 |
| Author of this report: | Peter Elias, Ritva Ellison, Rob Jones (Warwick, IER) |



Development of associated software for EurOccupations database

Summary

This note describes the development of a software tool which has the following functions:

- Allows users to interrogate the EurOccupations database using freeform natural language occupational descriptions
- Facilitates search of the EurOccupations database in seven languages
- Links results either to the EurOccupations extended list of occupations or to the key list of occupations
- Allows users to 'tailor' the software for national specificities (e.g. abbreviated words, synonyms, redundant words, etc.)

The software has been developed within the EurOccupations project from an existing software tool known as Cascot (Computer assisted structured coding of text). For further details see

<http://www2.warwick.ac.uk/fac/soc/ier/publications/software/cascot>

Background

Cascot was developed by the Warwick Institute for Employment Research (IER) via a research award from the UK Economic and Social Research Council. The software has been used primarily for two purposes – coding text to the UK Standard Occupational Classification and to NACE 2007. However, the software was designed to handle other classifications. Given that the EurOccupations extended list of occupations has been coded to ISCO'08 – the latest version of the International Standard Classification of Occupations – it was possible to adapt the software to enable it to code occupational information in each of the languages used within the EurOccupations database. Further development of the software was required for this purpose – particularly its capability to handle characters not found in the English language.

Cascot

To fully appreciate the flexibility of the software, users must first have some familiarity with its operation as a coding tool. For this purpose, a demonstration presentation has been prepared, showing the facilities it offers and its functionality. This presentation can be initiated by clicking on the link below

http://www2.warwick.ac.uk/fac/soc/ier/publications/software/cascot/Cascot_demonstration.ppt



Cascot for EurOccupations

Distribution of Cascot for EurOccupations

To distribute the software 'Cascot for EurOccupations' to the users a file package has been created. The package is called 'Cascot for EurOccupations V1.0.zip' and includes the following:

- Files which need to be installed onto a computer in order to use the Cascot for EurOccupations software;
- Installation instructions indicating the requirements for the computer, and showing how to install and activate the software;
- Two EurOccupations classifications in English and Dutch;
- Demonstration, presentation and tutorial files which are aimed at helping the user to learn how to use the software;
- Additional files which can be installed; these are aimed for advanced users who wish to experiment with creating new classifications or want to edit the existing ones.

The file package is delivered in association with this report and can be found on the project website www.euroccupations.org.

Functionality of Cascot for EurOccupations

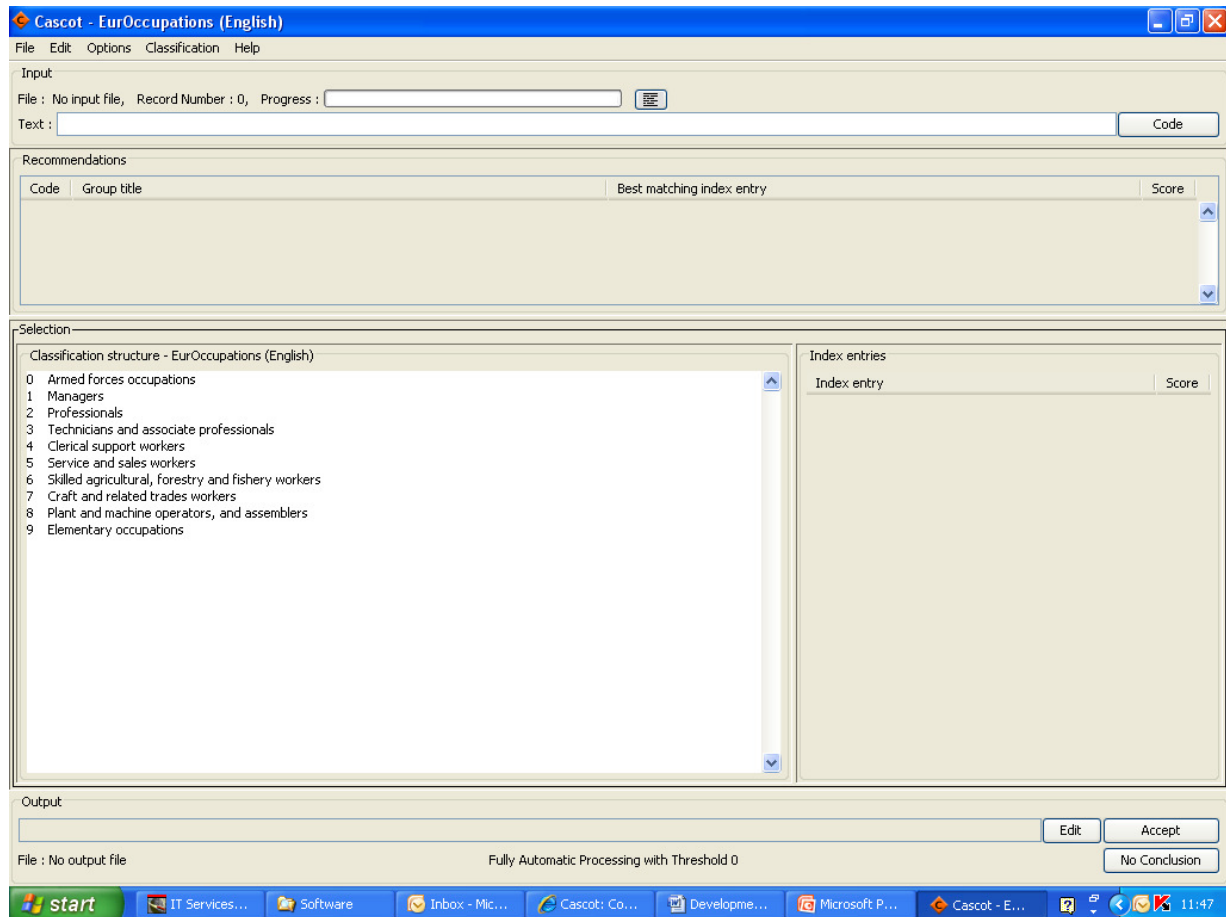
The functionality of the software is described in detail in the presentation, demonstration and tutorial files. A short description of the basic functionality is given below.

Figure 1 shows the first screen of the software after it has been started and the EurOccupations (English) classification has been opened. The classification is based on the ISCO'08 structure which can be seen in the *Classification structure* pane on the screen.

The two main basic functionalities of the software are

- manual coding of occupational texts
- automated coding of larger amounts of occupational texts

Figure 1: The start-up screen of EurOccupations (English) version

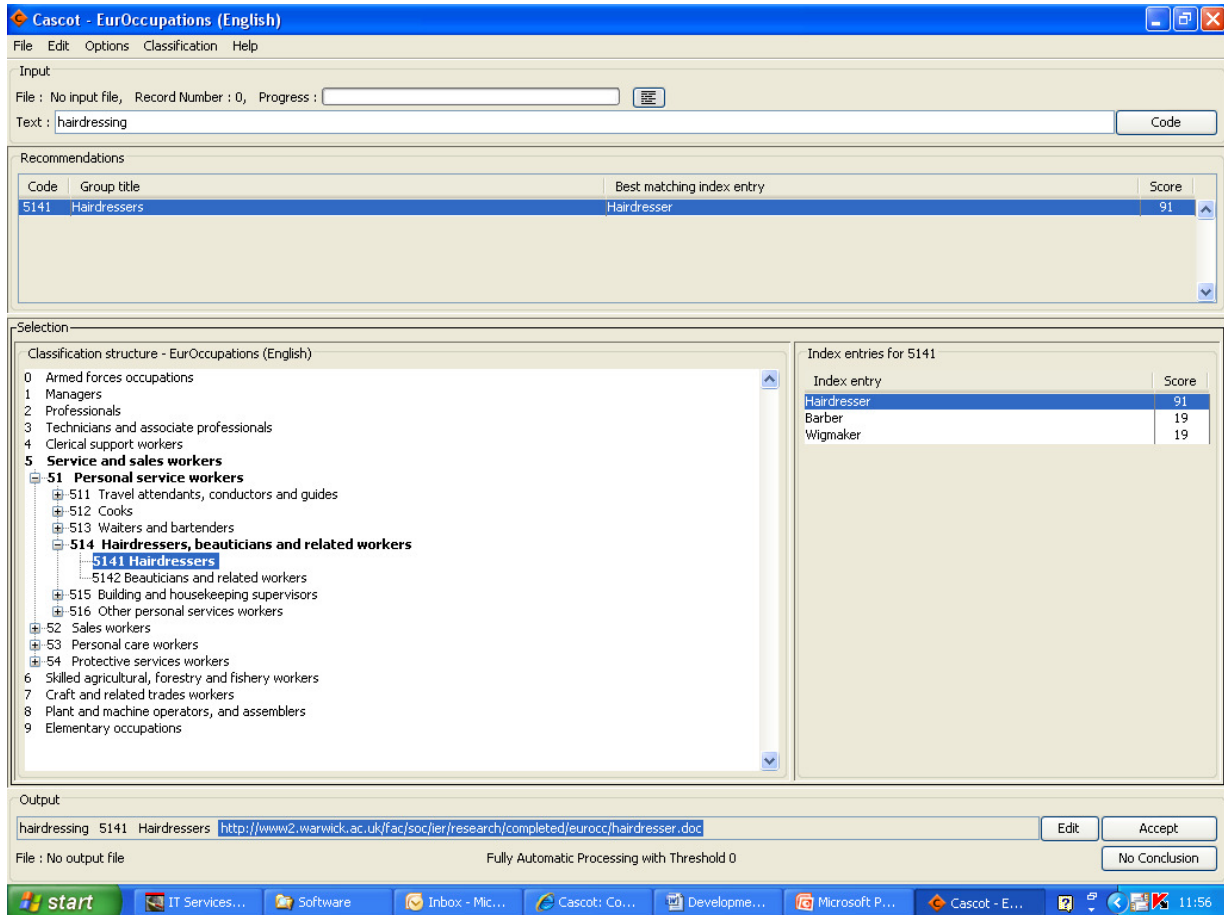


Manual coding

In the manual coding mode of operation the user can enter occupational text, either to find out the best matching associated EurOccupations occupational title and its ISCO'08 code, or to link to the EurOccupations documents showing the occupational analysis which was developed during the EurOccupations project. Figure 2 shows as an example the outcome after a user has entered manually the text 'hairdressing'.

As can be seen from Figure 2, the best matching EurOccupations title for this text string is 'Hairdresser' which has an ISCO'08 code 5141. All the resulting information is shown in the bottom pane which is called *Output*. Additionally, as Hairdresser is one of the EurOccupations project 150 key occupations, a link to a detailed document is shown in the *Output* pane (highlighted). By using the link with an Internet browser the relevant document can be easily retrieved.

Figure 2: An example of manual coding



The screenshot shows the Cascot software interface with the following components:

- Input:** Text field containing "hairdressing".
- Recommendations:** A table showing the best matching index entry for the input text.

| Code | Group title | Best matching index entry | Score |
|------|--------------|---------------------------|-------|
| 5141 | Hairdressers | Hairdresser | 91 |
- Selection:** A tree view of the classification structure. The selected path is:
 - 51 Personal service workers
 - 514 Hairdressers, beauticians and related workers
 - 5141 Hairdressers
- Index entries for 5141:** A table showing the index entries and their scores.

| Index entry | Score |
|-------------|-------|
| Hairdresser | 91 |
| Barber | 19 |
| Wigmaker | 19 |
- Output:** A text field showing the result: "hairdressing 5141 Hairdressers <http://www2.warwick.ac.uk/fac/soc/ier/research/completed/eurocc/hairdresser.doc>".

Automatic coding

For partial or fully automated coding, the user can have an unlimited number of occupational texts in a file which is fed through the software. The coded result is saved to an output file which can be used for different purposes. Table 1 shows an example of the output when 20 strings of English language occupational texts have been handled with the software.

On the IER website there is more information about the high volume usage of the software at http://www2.warwick.ac.uk/fac/soc/ier/publications/software/cascot/volume_use/

Table 1: An example of automatically coded text

| Original text | ISCO'08 | ISCO'08 title | Best matching EurOccupations title | Link to the relevant document |
|---------------------------------------|----------------|-------------------------------------|---|---|
| Account Manager | 3313 | Accounting associate professionals | Account manager for private clients | http://www.../extended_list_final.xls |
| Cashier with supervisory roles | 5230 | Cashiers and ticket clerks | Cashier | http://www.../cashier.doc |
| jobbing bricklayer | 7112 | Bricklayers and related workers | Bricklayer | http://www.../construction_bricklayer.doc |
| hgv driver | 8322 | Car, taxi and van drivers | Car driver | http://www.../extended_list_final.xls |
| Midwife | 2222 | Midwifery professionals | Midwifery professional | http://www.../midwifery_professional.xls |
| firefighter paramedic | 5411 | Fire-fighters | Fire fighter | http://www.../firefighter.doc |
| pharmacy storeman | 2262 | Pharmacists | Pharmacist, pharmacy manager | http://www.../extended_list_final.xls |
| Production planner | 4322 | Production clerks | Production planning clerk | http://www.../extended_list_final.xls |
| Photocopier Field Service Engineer | 7412 | Electrical mechanics and fitters | Refrigeration service engineer | http://www.../refrigeration_mechanic.doc |
| Freelance Photographer. | 3431 | Photographers | Advertisement photographer | http://www.../extended_list_final.xls |
| Payroll | 4313 | Payroll clerks | Payroll clerk | http://www.../payroll_clerk.doc |
| freight co-ordinator | 9333 | Freight handlers | Freight handler | http://www.../extended_list_final.xls |
| registered childminder | 5311 | Child care workers | Child-carer | http://www.../child_carer.xls |
| Industrial Engineer | 2141 | Industrial and production engineers | Industrial engineer | http://www.../extended_list_final.xls |
| Claims Assessor | 3315 | Valuers and loss assessors | Insurance claims examiner | http://www.../extended_list_final.xls |
| sales manager | 3322 | Commercial sales representatives | After sales manager | http://www.../extended_list_final.xls |
| IT Project Manager | 2511 | Systems analysts | IT project leader | http://www.../extended_list_final.xls |
| Taking and invoicing customer orders. | 4321 | Stock clerks | Order clerk | http://www.../extended_list_final.xls |
| Water scientist | 2114 | Geologists and geophysicists | Environmental scientist | http://www.../extended_list_final.xls |

| | | | | |
|------------------------|------|-------------|------------------------|---|
| Fixed Asset Accountant | 2411 | Accountants | Accountant (qualified) | http://www.../extended_list_final.xls |
|------------------------|------|-------------|------------------------|---|

Information about Cascot for EurOccupations

The general presentation about Cascot has 79 slides with examples that are related to the UK Standard Occupational Classification 2000. The presentation gives an overview of the software, its development and discusses occupational coding in general. It shows in detail the various parts of Cascot and explains the different modes of coding. This general presentation file 'Cascot_demonstration.ppt', is included in the distributed file package and is also available at

http://www2.warwick.ac.uk/fac/soc/ier/publications/software/cascot/Cascot_demonstration.ppt

In the distributed file package there is a 22-page Cascot tutorial, 'cascot_tutorial.pdf' with associated tutorial files. The tutorial is aimed at new users as it instructs and encourages the user to practice the different modes of coding. Again, the tutorial is based on the UK SOC2000 classification.

Finally, there are two demonstrations about Cascot for EurOccupations. They were presented in the final conference in Leuven in April 2009. The presentations discuss the adaptation of Cascot for EurOccupations and show in detail how Cascot for EurOccupations works in practice. The first presentation file has 48 slides and the second 73 slides. The files, 'Cascot_eurocc_demonstration_1.ppt' and 'Cascot_eurocc_demonstration_2.ppt', are included in the distributed file package and are also available at

http://www2.warwick.ac.uk/fac/soc/ier/research/completed/eurocc/cascot_eurocc_demonstration_1.ppt

and

http://www2.warwick.ac.uk/fac/soc/ier/research/completed/eurocc/cascot_eurocc_demonstration_2.ppt



Appendix Cascot EurOccupations demonstration