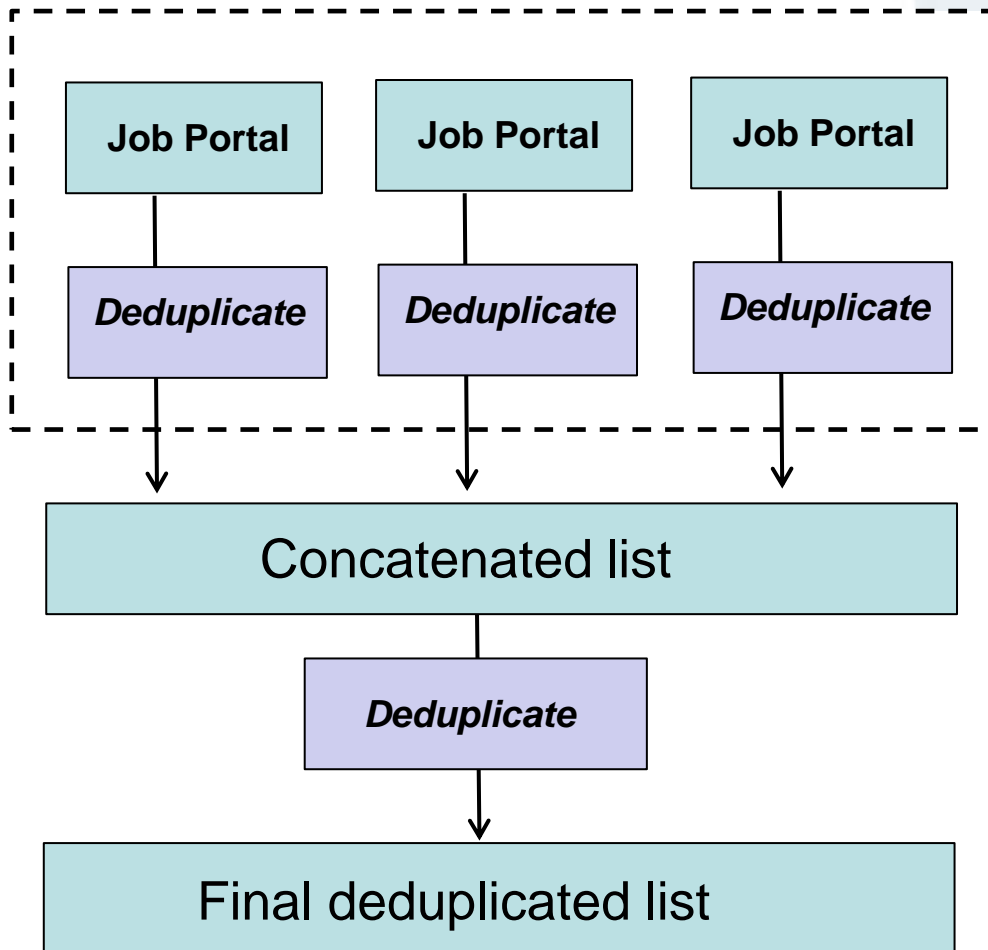


ESSNet Pilot:

**WP1 - Web Scraping for Job Vacancy
Statistics - UK update**

7th November 2016, Rome

Removing Duplicate Job Ads



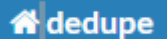
1. Create common variable list:

- Job_title
- Job_description
- Location_city
- Location_region
- Date_posted
- Enterprise name

2. Clean data:

e.g. ".NET Developer - Stoke-On-Trent - £35-£40K "

3. Run dedup to produce candidate matches

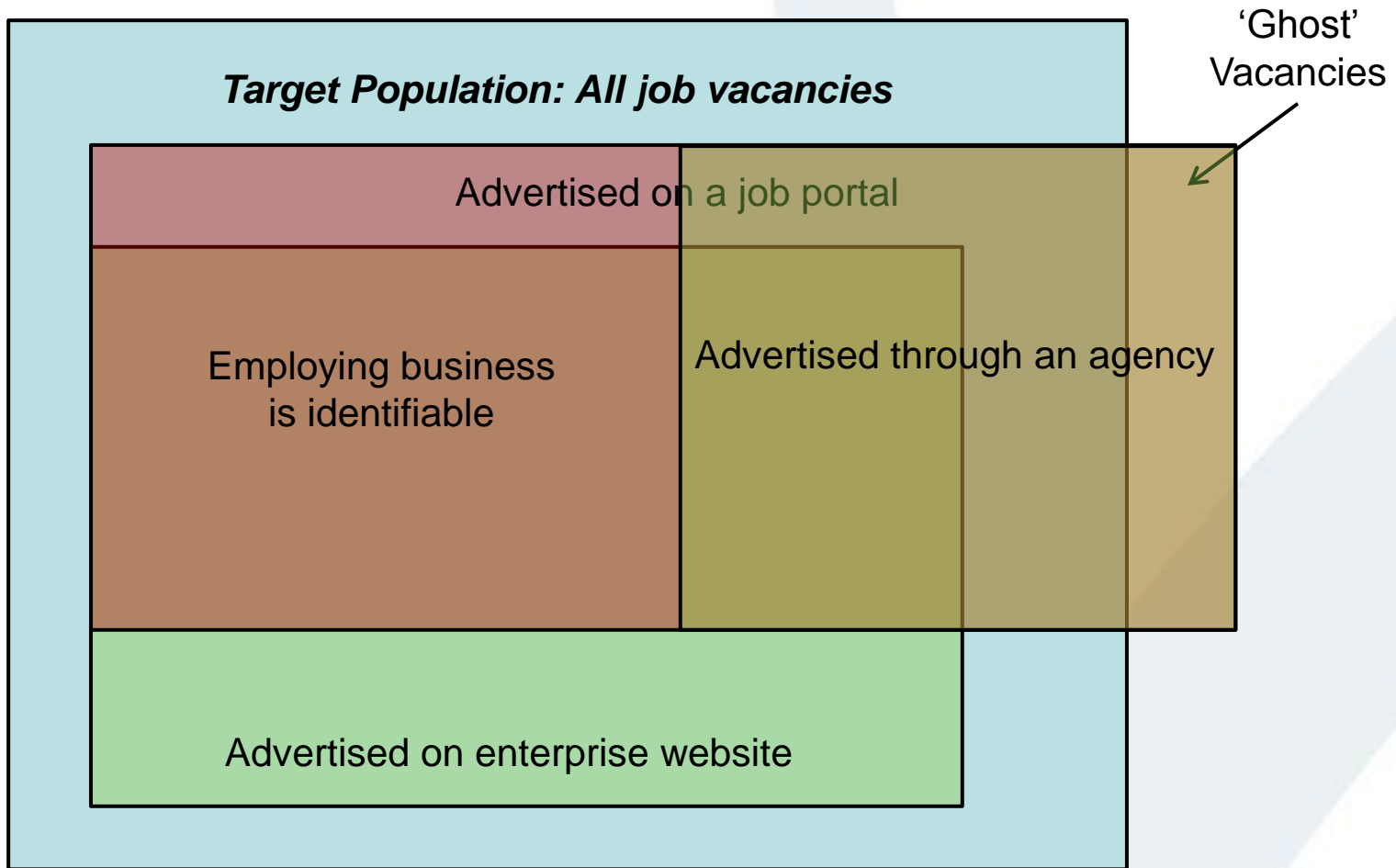


4. Active learning step (manual coding of > 100 records)

5. Rerun to automatically remove "duplicate" job ads

A lot of high quality training data needed to work effectively!

Conceptual model for measuring job vacancies from on-line sources



Current Workplan

- Matching job ad counts by advertising business from five portals to JV survey
- Focusing on 1300 largest reporting units (approx 33% of all JVs)
- About 25% can be matched easily
 - Need more BR data?
 - Manual matching of residuals?
- What about smaller enterprises?
 - Single location enterprises may be easier?

Data collection

careerJET

Job title, keywords or company name

Town, county or country (optional)

Job search > United Kingdom jobs > Company A

740 companies recruiting

A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | 1

A14 Integrated Delivery Team (12)

A1 Jobs Ltd (9)

A1 Personnel Employment (35)

A4 Scotland (12)

A & C Recruitment LTD (83)

A & D Recruitment Ltd (30)

A H I (6)

A Plant Hire Co Ltd (54)

AA (284)

AAP3 (63)

Aarca Recruitment Ltd (13)

Aaron Wallis Recruitment & Training Ltd (27)

Atom Recruitment (22)

AB-Inbev (7)

Abacus Consulting (9)

Abacus Professional Recruitment (224)

ABB (12)

Abbatt Property Services (21)

A1 Direct Recruitment (10)

A1 People (85)

A1 Personnel Ltd (96)

A & B RECRUITMENT (20)

A&D Recruitment (25)

A F Blakemore - Retail (8)

A&P Resources (9)

A Plus Teachers (24)

AAK (9)

AAP3 Recruitment LTD (61)

Aaron Wallis Recruitment and Training Limited (152)

Aaron Wallis Sales Recruitment (51)

AB Consultancy (6)

Abaco Systems Limited (13)

Abacus Employment Services (111)

Abatec Ltd (32)

ABBATT GROUP (42)

Abbey Mortimer (6)

CEDEFOP



- Pilot system for online vacancy analysis
- 4.2 million job vacancies (5 countries UK, Ireland, Germany, Italy, Czech Republic)
- May 2016: Training session
- June 2016: Access agreement to online analysis system
=> Initial assessment: Limited functionality
- August 2016: Agreement to supply underlying data:
 - Very large file delivered requiring bespoke database solution
 - not what we were expecting!
- Latest: CEDEFOP tendering to undertake further work....

CEDEFOP tender

Description of the procurement:

During the implementation of the contract, work will be assigned to the contractor according to order forms. The main tasks to be performed over the duration of the contract are as follows:

1. produce the following reports from a landscaping exercise undertaken in each Member State. The landscaping exercise will enable Cedefop to assess the representativeness of the data and enable proper interpretation of results:
 - a. reports focusing on hiring practices and the use of online platforms by employers in EU countries;
 - b. a report focusing on pan-European or worldwide specialised occupational portals (vacancies usually not published on a national portal due to the global character of the jobs). This report will serve as the basis for further assessment if such portals are to be potentially included in further developments;
 - c. a report analysing EURES — European job mobility portal. This report will assess the ability of EURES to act like the reference point for further analysis;
 - d. a synthesis report on findings from the landscaping exercise and implications for producing real-time labour market information on skills demand using online vacancies, the representativeness and comparability of the information across occupations and countries;
2. extend the current Cedefop prototype to a fully-fledged instrument for online vacancy gathering and analysing across all 28 EU Member States. In detail:
 - a. develop the tool for scraping (extracting information) from a set of pre-defined websites (national job boards) and process the information for further analysis and classifications of occupations and skills;
 - b. set up a database of sufficiently discriminatory information on generic and occupation-specific skills and other critical job requirements in occupations across countries. Although the job advertisements are usually driven by job titles the final analysis should be carried out according to the ISCO-08 occupations, use the most up to date ESCO classification and potentially refer to EQF and NQF;
 - c. develop a front-end interface for analysis and presentation of the results;
 - d. ensure the transfer of all elements to Cedefop, including all respective programs, algorithms, dictionaries, classifications, taxonomies, data, etc.;

Review of UK Standard Occupation Classification (SOC)

- UK SOC (UK version of ESCO)
- Last reviewed in 2010
- Public consultation supported by analysis of new occupations (including 2011 Census)
- Proposal is to use bulk job titles (and skills) scraped from job portals
- Benefits:
 - High volume and up-to-date data to inform SOC review
 - New job titles to enhance SOC coding frame
 - Duplication, text mining and coding methods could be developed and applied to WP1 ESSNet pilot
- SOC Pilot focusing on job titles in IT sector

Current Challenges

- Staffing
- API limitations
- Working across two environments
- Resources for doing supervised learning and matching
- Paperwork, meetings and conferences!

Looking ahead

- Data scientist recruited last week!
- Continue with work plan (plus SOC review)
- Re-engagement with CEDEFOP (and successful bid?)
- Engagement with job portal owners?
- Preparing for end of SGA-1 technical report