

# Big Data for Labour Market Intelligence

Програма з розбудови спроможності 2022 р.

**Модуль 1: Технічне навчання**

**Заняття 5**

Доступ до бази даних, моделі даних та основних функцій - Практичні рекомендації

# Система створення даних

## Туніс (04.2020-07.2022)

- ✓ 680 191 ОДВ - > 175 203 дедупліковані дані

## Україна (національний проект) (04.2020-07.2022)

- ✓ 2 571 655 ОДВ - > 1 304 262 дедупліковані дані

## Грузія (04.2021-07.2022)

- ✓ 129 271 ОДВ - 84 817 дедупліковані дані

## Єгипет (новий проект)

- ✓ 1 307 678 ОДВ– 391 701 дедупліковані дані

## Кенія (новий проект)

- ✓ (Збір даних розпочався у вересні 2022 р.)

# Збір та декодування даних про ринок праці

Дані про ринок праці в реальному часі дозволяють отримати інформацію, яку неможливо отримати з традиційних джерел



**Збір даних  
про ринок  
праці**

**Тегування та  
структурування  
даних**

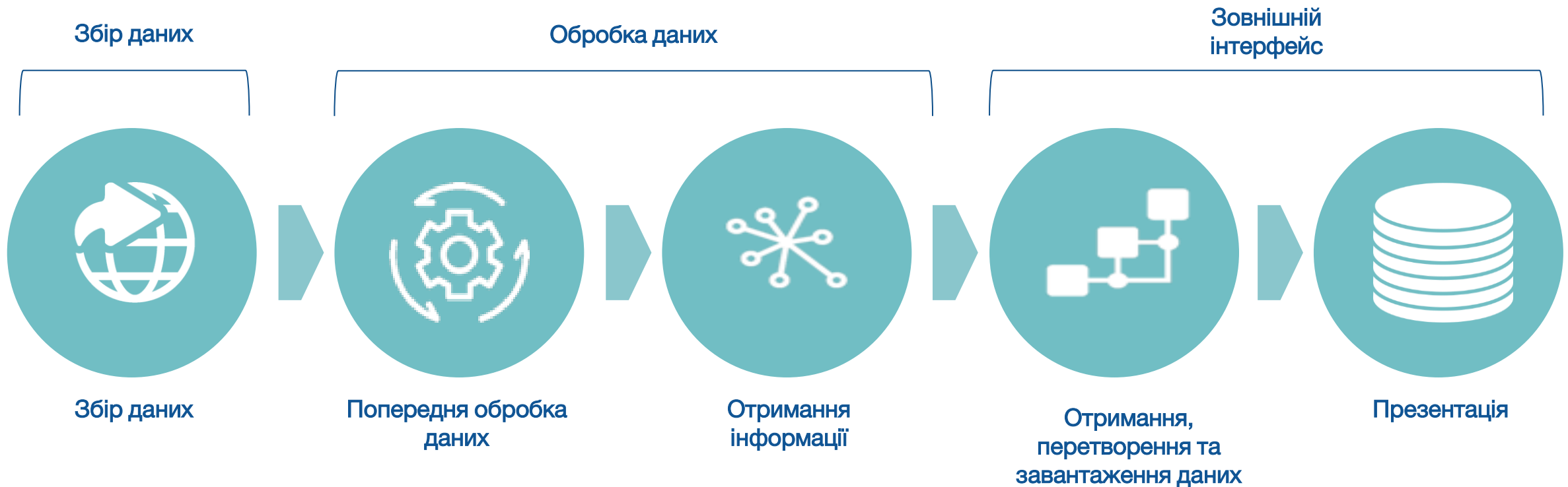
**Спільна  
мова**

Онтологія даних  
дозволяє робити  
співставлення

**Формулювання  
висновків**

Інформація про уміння, які  
користуються попитом, та  
характерні особливості кар'єрного  
росту з реального життя

# Загальний потік даних



# Отримання даних

- **Мета:**
  - Зібрати велику кількість неоднорідних даних з різних джерел непередбачуваної кількості
- **Виклики:**
  - Забезпечити повноту та послідовність даних
- **Підхід:**
  - Розробити підхід на основі різних методів (веб-сканування, скрейпінг, API-інтерфейс), що враховує особливості різних інтернет-сторінок: кількість вакансій, різні мови, структуру технологій, неінвазивний підхід та політичні угоди.
  - Попередження втрати джерел даних за допомогою резервування
  - Виявлення та збір метаданих для підвищення цінності інформації
- **Функції:**
  - Для того, щоб забезпечити управління процесом на етапі отримання даних, нами було розроблено систему моніторингу з інструментом планування та модулем, що генерує попередження

# Попередня обробка даних

- **Мета:**
  - **Забезпечити належні дані на етапі отримання інформації**
- **Виклики:**
  - **Вимірювання, моніторинг та покращення якості даних для забезпечення їхньої максимальної повноти, узгодженості, комплексності, актуальності та періодичності**
- **Підхід:**
  - **Розробити багатоетапний процес, що включає такі етапи:**
    - **Визначення вакансій: аналіз інтернет-сторінок для відбору лише контенту, що стосується вакансій**
    - **Дедуплікація: визначити дубльовані оголошення про роботу для отримання єдиної вакансії**
    - **Визначення дати: визначити дату публікації оголошення та дату закінчення його терміну за допомогою описового аналізу вакансій**
    - **Тривалість оголошення про роботу: метод визначення дати закінчення терміну оголошення, якщо такі дані відсутні**
- **Функції:**
  - **Забезпечення якості даних на всіх етапах обробки даних**

# Класифікація даних

- **Мета:**
  - Отримати з даних інформацію та структурувати її для презентації
- **Виклики:**
  - Обробка великого об'єму неоднорідних даних, викладених різними мовами
- **Підхід:**
  - Розробка гнучкої структури, залежної від мови, адаптованої для різних інформаційних компонентів. Деякі пов'язані з цим виклики:
    - Класифікація компоненту «Професія»: поєднання таких методів, як машинне навчання, тематичне моделювання і неконтрольоване навчання
    - Класифікація компоненту «Вміння»: поєднання різних методів, таких як аналіз текстів на основі корпусного підходу або подібності з використанням бази знань
- **Компоненти:**
  - Забезпечити отримання пояснювальної інформації, методів класифікації протоколювання та відповідні функції

# Сфера презентації



Презентація

Можливості дослідницького аналізу даних

Публікація й презентація даних та співробітництво

Самостійний аналіз даних та інтелектуальна обробка економічних даних

Вбудована аналітика з додатковими можливостями

Дослідження на основі візуалізації

Системи інтерпретації

Аналіз просторового розташування

Визначення та дослідження графіків

Розповідь про дані



# Визначення користувачів та схем навігації



Фахівці по роботі з даними та  
аналітики



Органи управління  
та  
корпоративні користувачі

# Фахівці по роботі з даними та аналітики



Можливості  
дослідницького  
аналізу даних



Публікація й  
презентація даних  
та співробітництво



Інтеграція  
машинного  
навчання



Вбудована  
аналітика з  
додатковими  
МОЖЛИВОСТЯМИ

# Органи управління та корпоративні користувачі



Самостійний аналіз даних  
та інтелектуальна обробка  
економічних даних



Дослідження на  
основі візуалізації

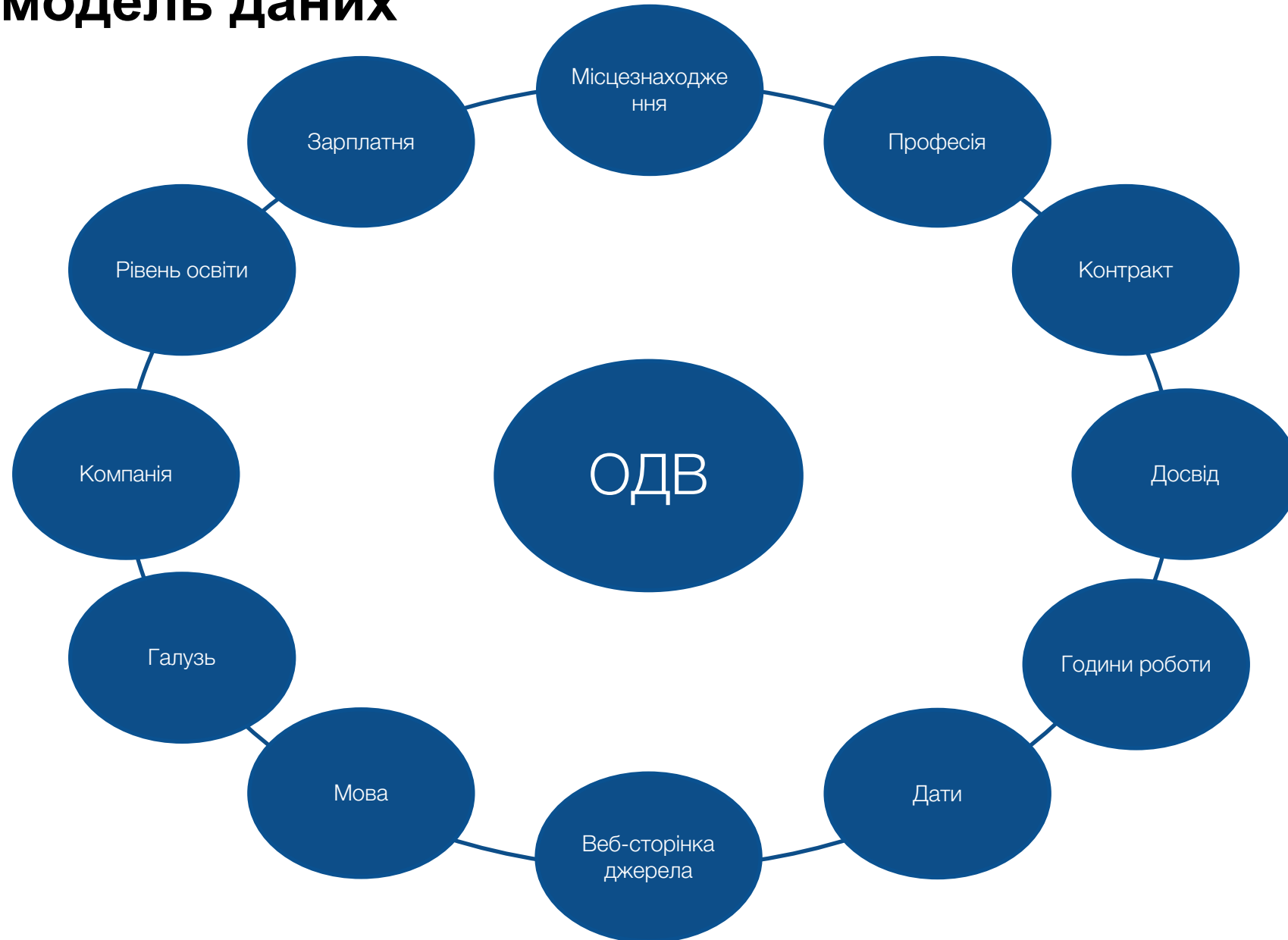


Контрольоване  
дослідження даних



Розповідь про дані

# (Базова) модель даних

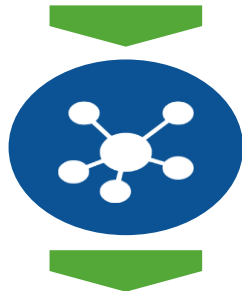


# Галузь

Посилання: 990-NHSE8576N  
Галузь: Охорона здоров'я  
Зарплатня: 56 665 – 69 168 на рік  
Місце: Лідс

NHS England є провідною установою національної системи охорони здоров'я (NHS) в Англії. Ми визначаємо пріоритети та напрямки розвитку NHS та заохочуємо обговорення на національному рівні питань охорони здоров'я та медобслуговування. Ми прагнемо того, щоб кожна людина мала більший контроль над своїм здоров'ям та добробутом і отримувала підтримку для того, щоб жити довше та щасливіше завдяки медичним послугам високої якості, що базуються на співчутті, інклюзивності та постійному удосконаленні...

В ОДВ є структуроване поле «Класифікатор NACE» (значення поля - «охорона здоров'я», що відповідає онтології)



**Класифікаційне значення поля:**  
**86 - діяльність в галузі охорони здоров'я**

## Посилання



# Контракт

Працівник - Ейлсбері  
Контракт: Тимчасовий (3 місяці)  
Зарплатня: £10 на годину

Шукаємо працюючого співробітника в районі Ейлсбері.  
Ви будете працювати в центральному закладі. Обов'язки включатимуть прибирання та допомогу відвідувачам. До закладу можна доїхати громадським транспортом. Також є парковка, якщо...

Посилання

Постійний  
Само-зайнятість  
Тимчасовий

Класифікація,  
визначена  
користувачем

В ОДВ є структуроване поле  
«Контракт» (значення поля -  
«тимчасовий», що відповідає  
онтології)



Класифікаційне значення поля:  
Тимчасовий

# Години роботи

Посилання: 184-SS.GEN.38

Відділення: Деменція

Місцезнаходження: Брекен Хауз, Чард

Команда допомоги в сфері психічного здоров'я для людей похилого віку міста Чард, бажає взяти на роботу **на часткову зайнятість** медсестру 5 категорії, що спеціалізується в галузі психічного здоров'я, для надання послуг з оцінки пам'яті та допомоги в роботі амбулаторії.

В якості співробітника команди ви будете тісно співпрацювати з середнім медичним персоналом та підрозділом реабілітації району, а також терапевтами, соціальними працівниками, підрозділом гострих станів та волонтерами...

Години роботи не вказані в якості структурованого поля в ОДВ, але у тексті йдеться про години роботи («часткова зайнятість»), що відповідає онтології



Класифікаційне значення поля:  
Часткова зайнятість

Посилання

Повна  
зайнятість  
Часткова  
зайнятість

Класифікація,  
визначена  
користувачем

# Рівень освіти

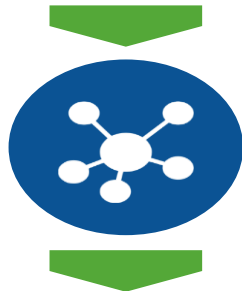
Посада: керівник групи з питань рухомого складу  
Місцезнаходження: Південний Лондон  
Зарплата: приблизно £47 500  
Вимоги до рівня освіти: ступінь молодшого спеціаліста  
Досвід: до 1 року

Мета - скеровувати щоденну роботу з вчасного надання рухомого складу з дотриманням відповідних стандартів технічного обслуговування та створення умов для забезпечення руху пасажирів. Посадові обов'язки, зокрема, включають: щоденне надання транспортних засобів для роботи, що забезпечується надійно та послідовно. Входить до складу керівної команди, що забезпечує ефективне технічне обслуговування з мінімальними затратами...

Посилання

ISCED  
2011

В ОДВ є структуроване поле  
«Вимоги до рівня освіти»  
(значення поля - «ступінь  
молодшого спеціаліста», що  
відповідає альтернативній назві  
онтології)



**Класифікаційне значення поля:  
Бакалавр або еквівалентна ступінь**



# Зарплатня

Посада: керівник групи з питань рухомого складу  
Місце: Південний Лондон  
Зарплатня: приблизно £47 500  
Вимоги до рівня освіти: ступінь молодшого спеціаліста  
Досвід: до 1 року

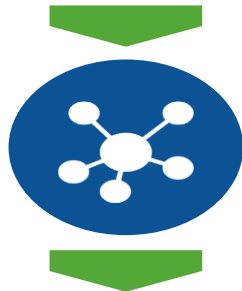
Завдання - скеровувати щоденну роботу з вчасного надання рухомого складу з дотриманням відповідних стандартів технічного обслуговування та створення умов для забезпечення руху пасажирів. Посадові обов'язки, зокрема, включають: щоденне надання транспортних засобів для роботи, що забезпечується надійно та послідовно. Входить до складу керівної команди, що забезпечує ефективне технічне обслуговування з мінімальними затратами...

Посилання

13 рівнів

Класифікація,  
визначена  
користувачем

В ОДВ є структуроване поле «Зарплатня» (значення поля - «£47 500 на рік», перераховані в Євро)



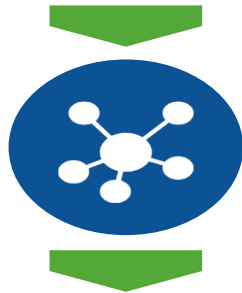
Класифікаційне значення поля:  
48 000 - 54 000 Євро на рік

# Досвід

Посада: керівник групи з питань рухомого складу  
Місце: Південний Лондон  
Зарплатня: приблизно £47 500  
Вимоги до рівня освіти: ступінь молодшого спеціаліста  
Досвід: до 1 року

Мета - скеровувати щоденну роботу з вчасного надання рухомого складу з дотриманням відповідних стандартів технічного обслуговування та створення умов для забезпечення руху пасажирів. Посадові обов'язки, зокрема, включають: щоденне надання транспортних засобів для роботи, що забезпечується надійно та послідовно. Входить до складу керівної команди, що забезпечує ефективне технічне обслуговування з мінімальними затратами...

В ОДВ є структуроване поле «Досвід» (значення поля - «менше 1 року», що відповідає онтології)



Класифікаційне значення поля:  
Менше 1 року

Посилання

8 рівнів

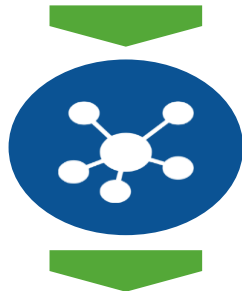
Класифікація,  
визначена  
користувачем

# Місцезнаходження

Посилання: 990-NHSE8576N  
Галузь: охорона здоров'я  
Зарплатня: 56 665 – 69 168 на рік  
Місце: Лідс

NHS England є провідною установою національної системи охорони здоров'я (NHS) в Англії. Ми визначаємо пріоритети та напрямки розвитку NHS та заохочуємо обговорення на національному рівні питань охорони здоров'я та медобслуговування. Ми прагнемо того, щоб кожна людина мала більший контроль над своїм здоров'ям та добробутом і отримувала підтримку для того, щоб жити довше та щасливіше завдяки медичним послугам високої якості, що базуються на співчутті, інклюзивності та постійному удосконаленні...

В ОДВ є структуроване поле «Місцезнаходження» (значення поля - «Лідс», що відповідає онтології)



Класифікаційне значення поля:  
Лідс

Посилання



# Професія

Фахівець по роботі з системою «Unix»

На цій посаді ви матимете наступні обов'язки:

- o Встановлювати та забезпечувати роботу операційної системи сервера та її сервісів, а також ПЗ для управління системою
- o Налаштовувати конфігурацію операційної системи
- o Здійснювати управління системами файлів та чергою на друк
- o Здійснювати моніторинг та забезпечувати підтримку журнальних файлів операційної системи
- o Давати рекомендації стосовно оновлення операційної системи та зміни її конфігурації

Алгоритм машинного навчання встановив зв'язок з вірною професією, яка відсутня в онтології



Класифікаційне значення поля:  
2522 - Системний адміністратор

Посилання  
(4<sup>й</sup> рівень ESCO)

ESCO v1.1  
ISCO08  
(4<sup>о</sup>-DIGIT)

# Уміння

Ви - досвідчений адміністратор в пошуку свого наступного контракту в районі Бристоля?  
Наш клієнт - спеціаліст з технічного обслуговування великого об'єкту, в якого є можливість негайно взяти на роботу співробітника на посаду адміністратора підрозділу на початкових умовах тимчасового найму

Обраний кандидат виконуватиме ряд **адміністративних** завдань, що включають **відповіді на дзвінки, комунікацію з підрядниками та виставлення рахунків**.

Обов'язки:

- **Використовувати ІТ-систему** для надання адміністративних послуг з підготовки, обробки та відбору кошторисів, пропозицій та тендерів
- **Замовлення** товарів, матеріалів та послуг для виконання умов контрактів
- **Внутрішня та зовнішня комунікація**, ведення записів або створення звітів за потреби
- Забезпечення правильності та вчасності всіх даних, документів та звітів по контрактах
- Підтримка керівництва **в забезпеченні потреб ділової активності**
- Професійно та ефективно **вирішувати питання, пов'язані з запитами або повідомленнями клієнтів/споживачів**.

Вимоги:

- Впевнені навички в галузі ІТ, **компетентність у використанні MS Office**
- Відмінні **навички комунікації** як усні, так і письмові
- Відмінний організатор з підтвердженими **уміннями в сфері управління часом**
- ....

Посилання

ESCO V1.1  
+  
Визначена  
користувачем

# Набір даних



# Набір даних - Модель та сховище даних

## Ключові слова

- Денормалізовані таблиці
- Стовбчикове хранилище даних

# Модель набору даних

## Документ FT

1 рядок для кожного з параметрів:

- General\_ID [?] Key (ключ загального ідентифікатора)
- ОДВ
- Джерело
- Місцезнаходження

Чому? Тому що для кожного ОДВ ми можемо визначити вакансію з декількома локаціями

Наприклад, «розробник ПЗ у Лондоні/Ліверпулі»

2 таблиці

## Аналіз умінь FT

1 рядок для кожного з параметрів:

- General\_ID [?] Key (ключ загального ідентифікатора)
- Джерело
- ОДВ
- Місцезнаходження
- Уміння

Чому? Тому що для кожного ОДВ ми можемо визначити декілька вмінь.

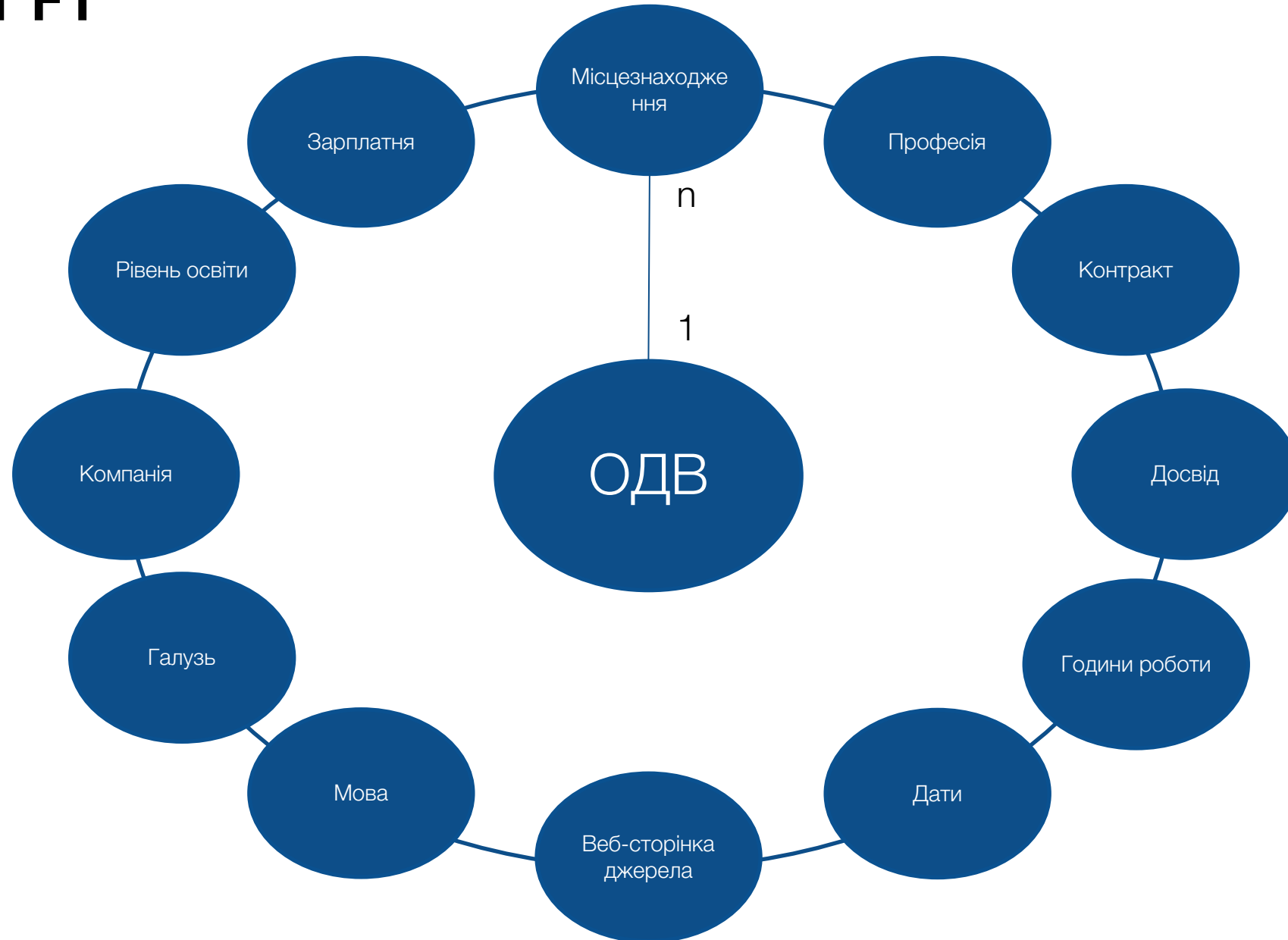
Наприклад, «розробник ПЗ у Лондоні/Ліверпулі з підходом, орієнтованим на потреби клієнтів, що володіє англійською та вміє справлятися зі стресом»



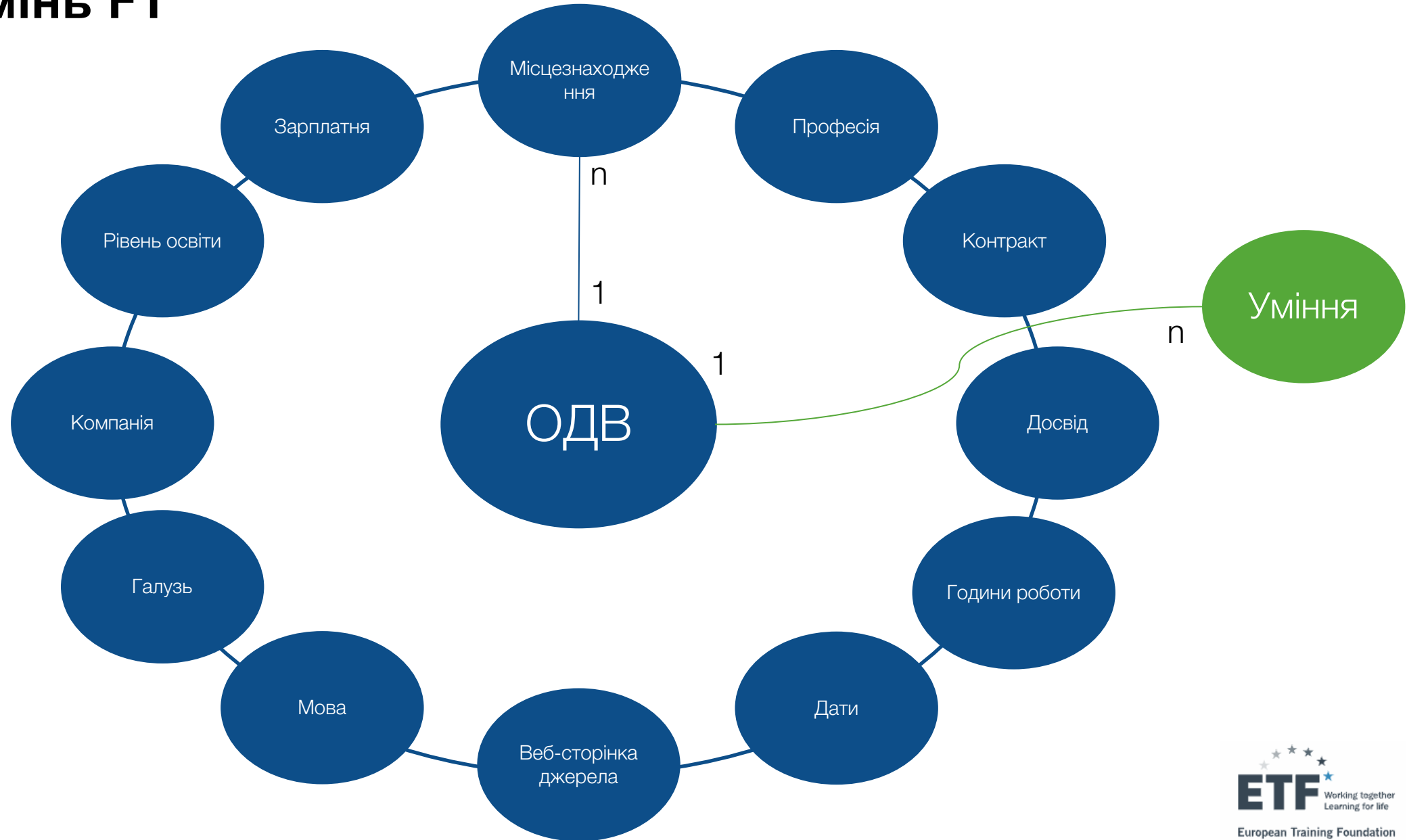
# Модель набору даних

Отже, для того, щоб отримати кількість вакансій, необхідно розрахувати унікальний показник на основі загального ідентифікатора

# Документ FT



# Аналіз умінь FT



# Інструменти

