Big Data for Labour Market Intelligence

Програма з розбудови спроможності 2022 р.

Модуль 1: Технічне навчання

Заняття 9

Використання великих даних для аналізу динаміки професій та вмінь на ринку праці

Поєднання даних з різних джерел

Доповідач: Мауро Пелуккі

29.11.2022

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

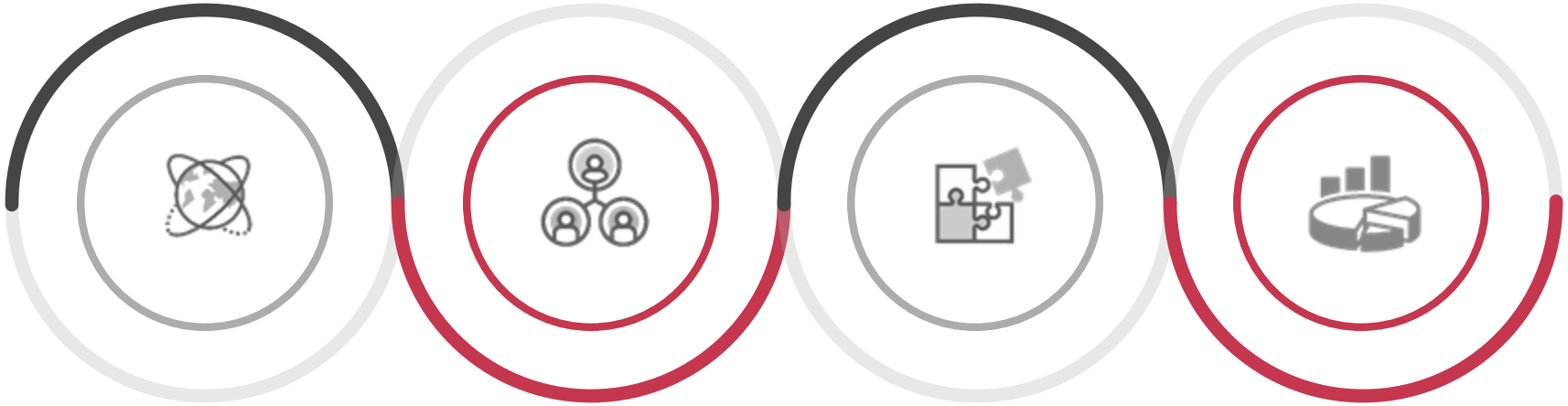
Кенія (новий проект)

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

ETF Working together Learning for life
European Training Foundation

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

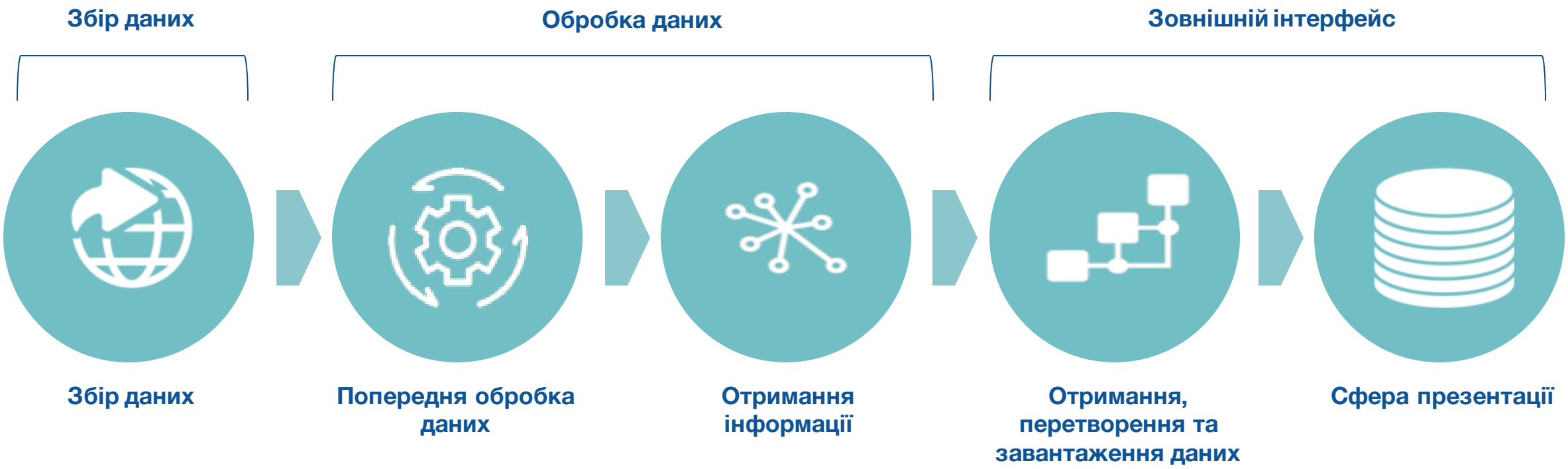
Онтологія даних дозволяє робити співставлення

Формулювання висновків

Інформація про вміння, які користуються попитом, та характерні особливості карьєрного росту з реального життя

ETF Working together Learning for life
European Training Foundation

Загальний потік даних



European Training Foundation

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
- **Компоненти:**
  - Для того, щоб забезпечити управління процесом на етапі отримання даних, нами було розроблено систему моніторингу з інструментом планування та модулем, що генерує попередження

ETF Working together Learning for life
European Training Foundation

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:**
    - **Виявлення вакансій: аналіз інтернет-сторінок для відбору лише контенту, що стосується вакансій**
    - **Дедуплікація: визначити дупліковані оголошення про роботу для отримання єдиної вакансії**
    - **Визначення даних: визначити дати публікації оголошення та дати закінчення його дії за допомогою описового аналізу вакансій**
    - **Тривалість вакансії: метод визначення дати закінчення терміну дії оголошення, якщо такі дані відсутні**
- **Функції**
  - **Забезпечення якості даних на всіх етапах обробки даних**

ETF Working together Learning for life
European Training Foundation

Класифікація даних

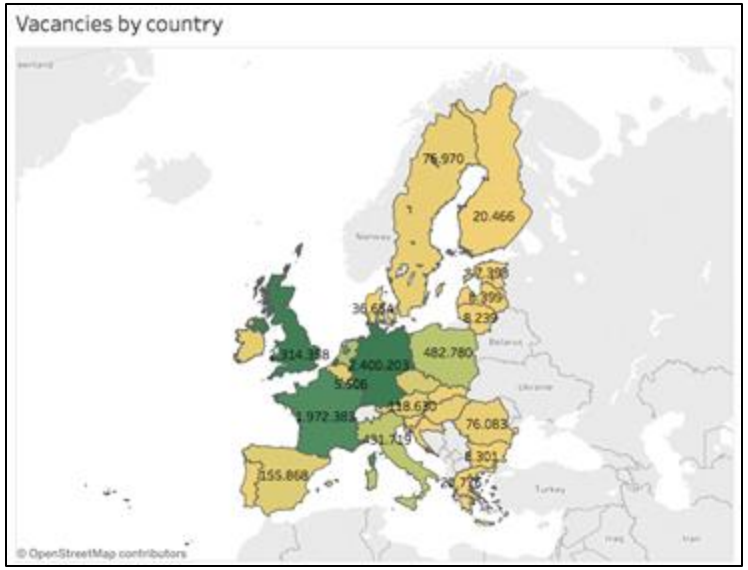
- **Мета:**
  - **Отримати з даних інформацію та структурувати її для презентації даних**
- **Виклики:**
  - **Обробка великого об’єму неоднорідних даних, викладених різними мовами**
- **Підхід:**
  - **Розробка гнучкої структури, залежної від мови, адаптованої для різних компонентів інформації. Деякі пов’язані з цим виклики:**
    - **Класифікація компоненту «Професія»: поєднання таких методів, як машинне навчання, тематичне моделювання і неконтрольоване навчання**
    - **Класифікація компоненту «Уміння»: 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**

ETF Working together Learning for life
European Training Foundation

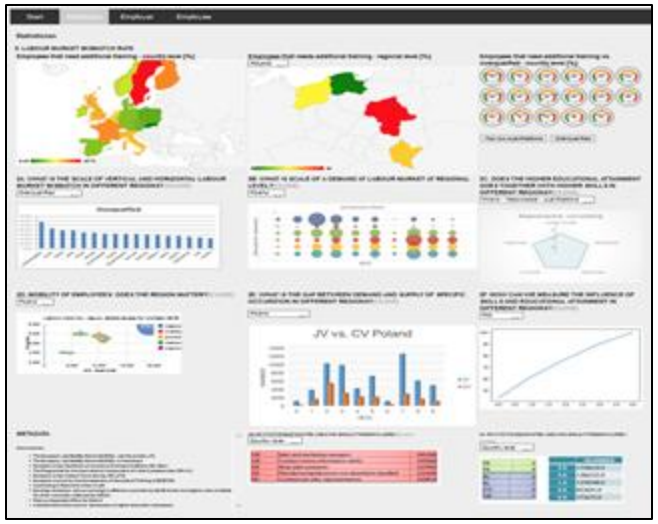
Презентація

- **Мета:**
  - Надати дані в аналітичному середовищі для забезпечення потреб користувачів у здійсненні різноманітних видів аналізу
- **Виклики:**
  - Персоналізований доступ до даних для різних категорій зацікавлених осіб
- **Підхід:**
  - Розробити різні шляхи навігації для різних користувачів:
    - **Фахівці по роботі з даними/аналітики:** доступ до даних через середовище «Discovery Lab», яке задовольнить їхні потреби щодо вільного аналізу даних.
    - **Органи управління:** доступ до інформації через розповідь про дані та динамічні інформаційні панелі.

ETF Working together Learning for life
European Training Foundation



Інформаційна панель з можливістю навігації за допомогою наведення курсора і натискання



Портал ринку праці
(Хакатон з великих даних 2017 р. - команда Польші)

Мапа (Хакатон з великих даних 2017 р. - команда Франції)

Інформаційна панель (Хакатон з великих даних 2017 р. - команда Норвегії)

ETF Working together Learning for life
European Training Foundation

Лабораторне заняття
Виявлена порівняльна перевага

Lightcast

Виявлена порівняльна перевага (RCA)

Given a set of occupations $\bar{O} = \{o_k, k = 1, \ldots, m\}$, a set of skills $\bar{S} = \{s_j, j = 1, \ldots, p\}$, and a matrix $\underset{m \times p}{M}$ that contains a value of sf for each pair of occupations $o_k \in \bar{O}$ and skills $s_j \in \bar{S}$, rca for o_i and s_l is defined as:

$$rca(o_i, s_l) = \frac{sf(o_i, s_l) / \sum_{j=1}^{p} sf(o_i, s_j)}{\sum_{k=1}^{m} sf(o_k, s_l) / \sum_{k=1}^{m} \sum_{j=1}^{p} sf(o_k, s_j)}$$

ETF Working together Learning for life
European Training Foundation

Подібність професій

Із блокноту:

- Ми оберемо країну
- Розрахуємо RCA для цієї країни
- І побачимо, як створити зв’язки між професіями

Cmd 1

RCA

In this notebook we will see the computetion of RCA and now use it to calculate a KPI about the distance from occupations.

Thanks to Simone Perego for the Notebook

Cmd 2

```
%md
select idcountry, idesco_level_4, esco_level_4, idescoskill_level_3, escoskill_level_3, count(distinct general_id) as num_ojv from ft_skill_analysis_en  where idesco_level_4 in ('1330', '2511', '2512', '2513', '2514', '2519', '2521', '2522', '2523', '2529', '3511', '3512', '3513', '3514', '3521', '3522') and year_grab_date = 2020 and idcountry in ('IT', 'DE', 'FR', 'ES', 'IE', 'CZ') and idescoskill_level_1 <> 'No digital'
group by idcountry, idesco_level_4, esco_level_4, idescoskill_level_3, escoskill_level_3
```

Cmd 3

```
%sql
select * from default.data_rca_csv limit 10
```

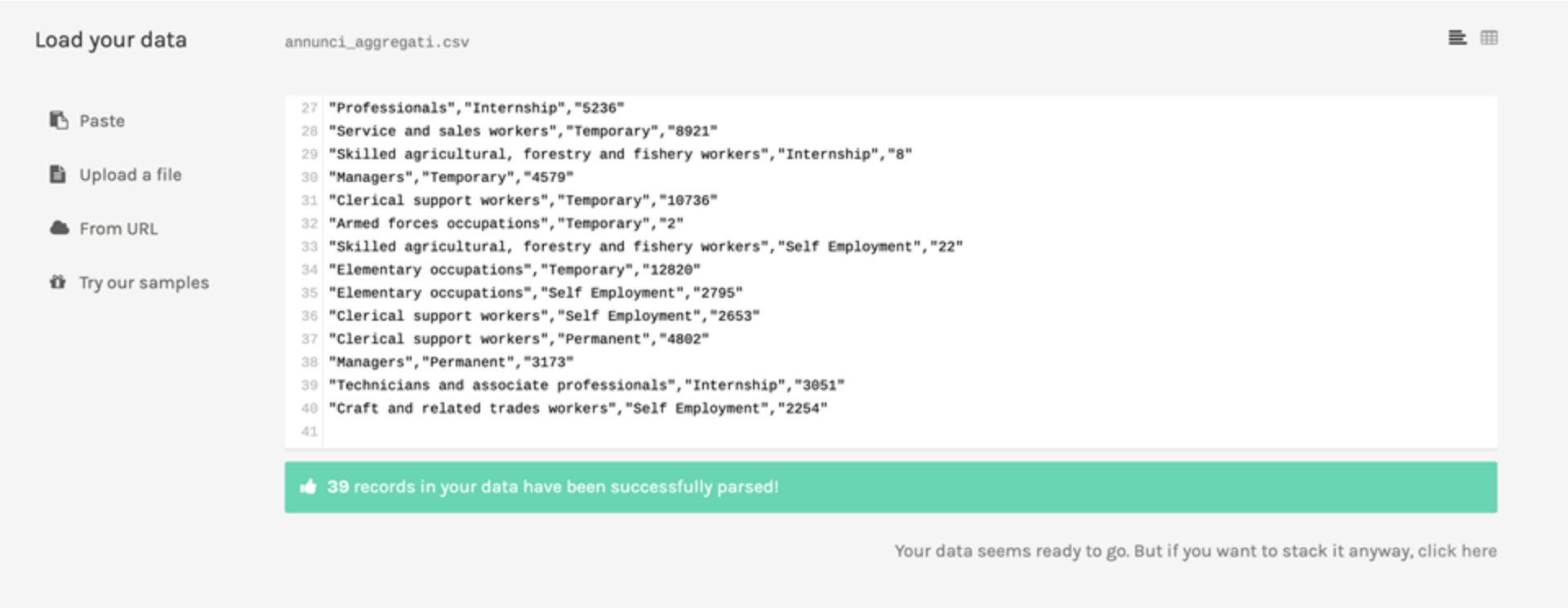
(1) Spark Jobs

	idcountry	idesco_level_4	esco_level_4	idescoskill_level_3	escoskill_level_3	num_ojv
1	FR	1330	Information and communications technology service managers	STACK_157	elasticsearch	112
2	CZ	2511	Systems analysts	ESCOv1_10465	use digital device operating systems	116
3	ES	3512	Information and communications technology user support technicians	ESCOv1_3737	manage system security	272
4	DE	3513	Computer network and systems technicians	ESCOv1_529	cloud technologies	6469
5	FR	1330	Information and communications technology service managers	STACK_40	eclipse	65
6	DE	2523	Computer network professionals	ESCOv1_4128	quality standards	1980

Showing all 10 rows.

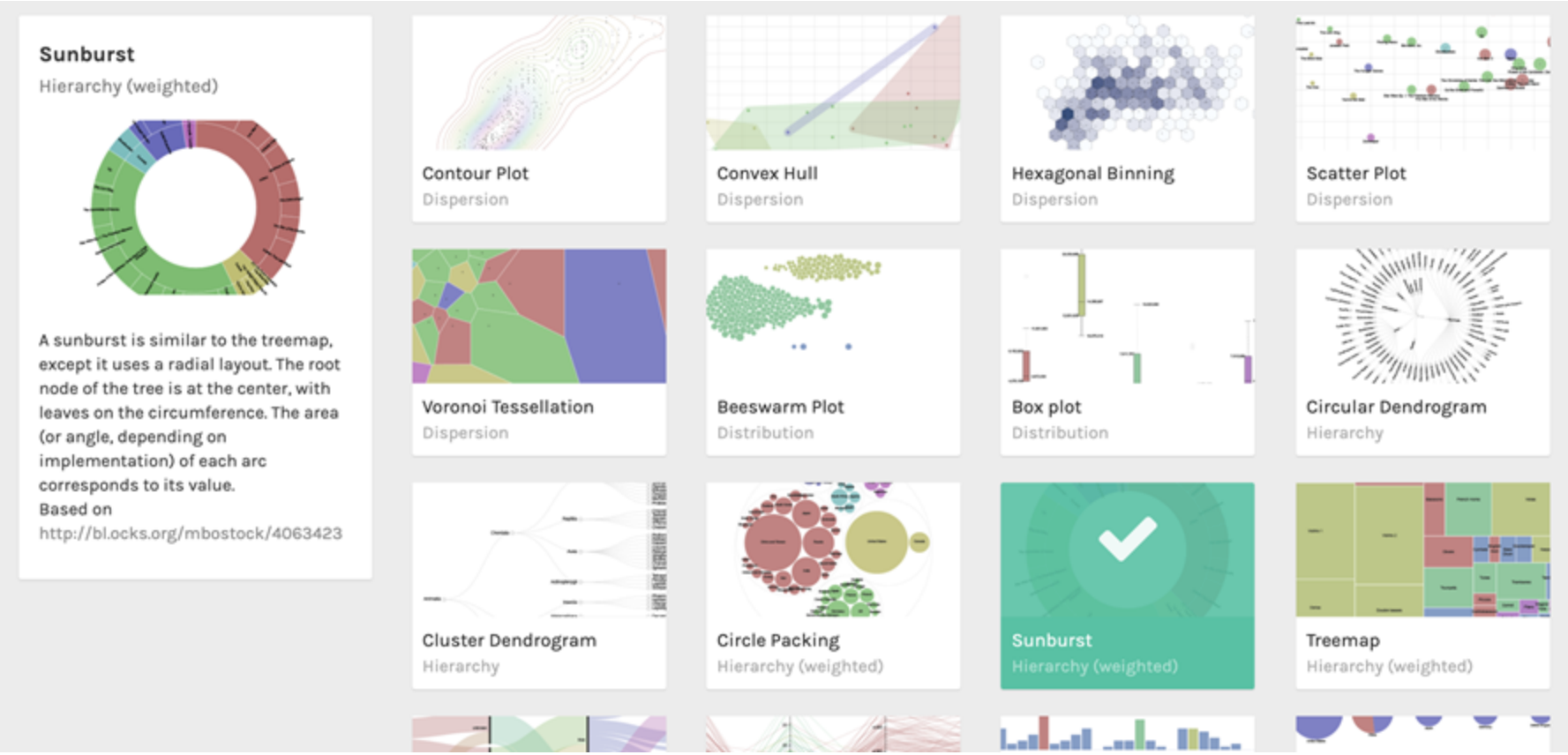
rawdata.io

https://app.rawgraphs.io/



rawdata.io

https://app.rawgraphs.io/

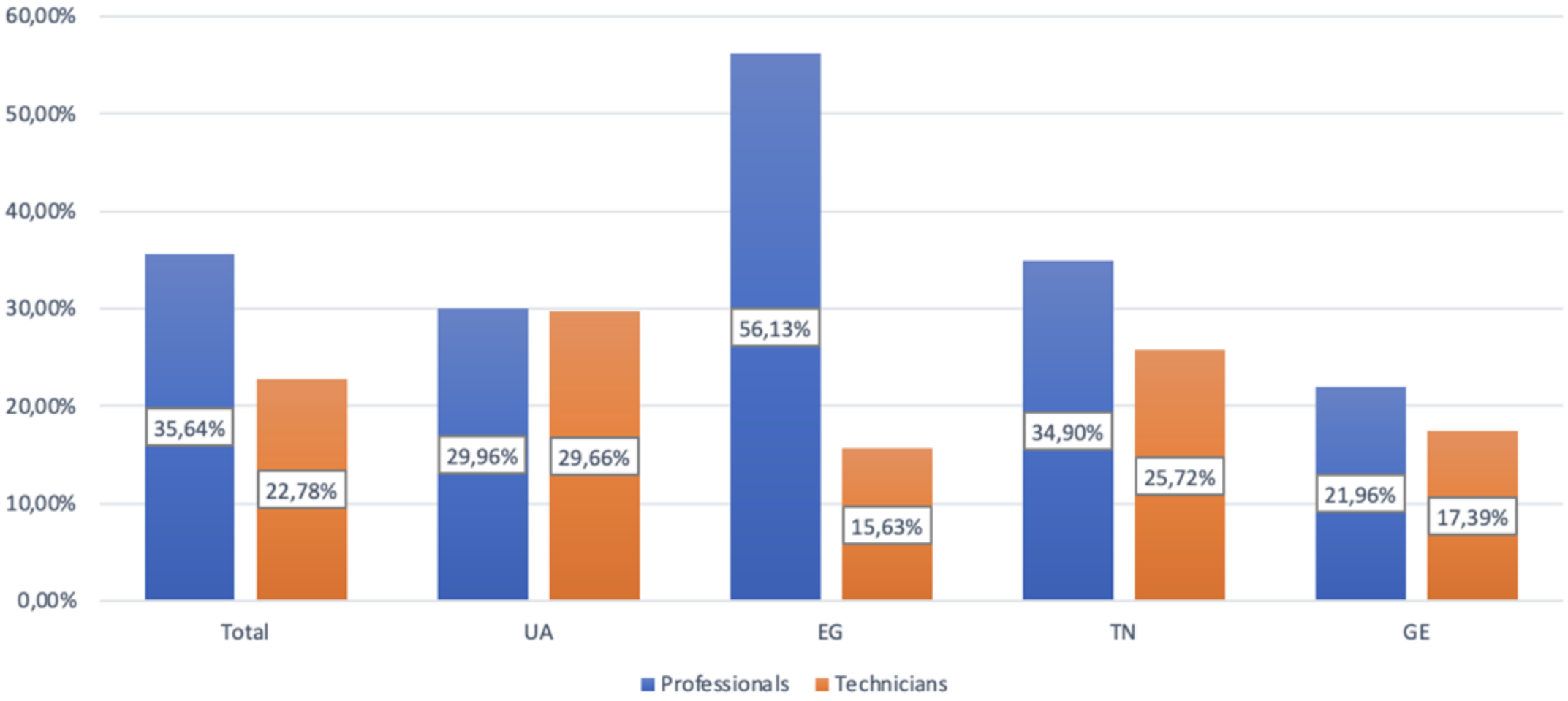


ETF Working together Learning for life
European Training Foundation

Уміння та професії

Lightcast

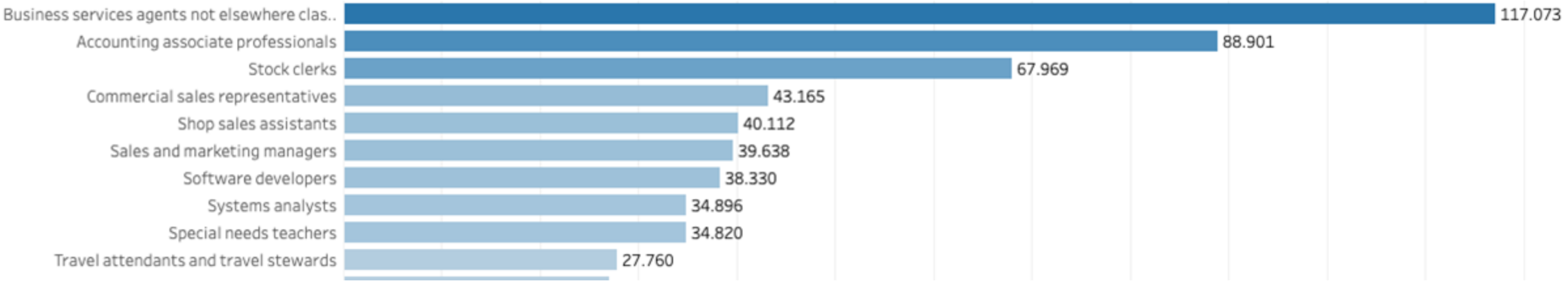
Важливість фахівців



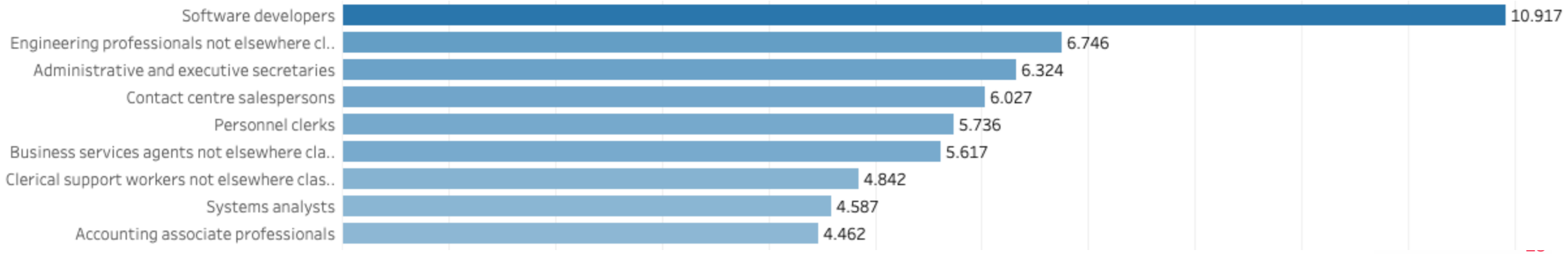
Джерело: Набір даних ЄФО - 2020 р.- вересень 2022 р.

Детальні дані про професії

Україна - Набір даних ЄФО - Унікальні оголошення про роботу 2020-2022 р.р.



Туніс - Набір даних ЄФО - Унікальні оголошення про роботу 2020-2022 р.р.



Уміння та ESCO

Класифікація «ESCO» дозволяє згрупувати уміння за категоріями. Основні відмінності між професійно-технічними та комунікативними уміннями:

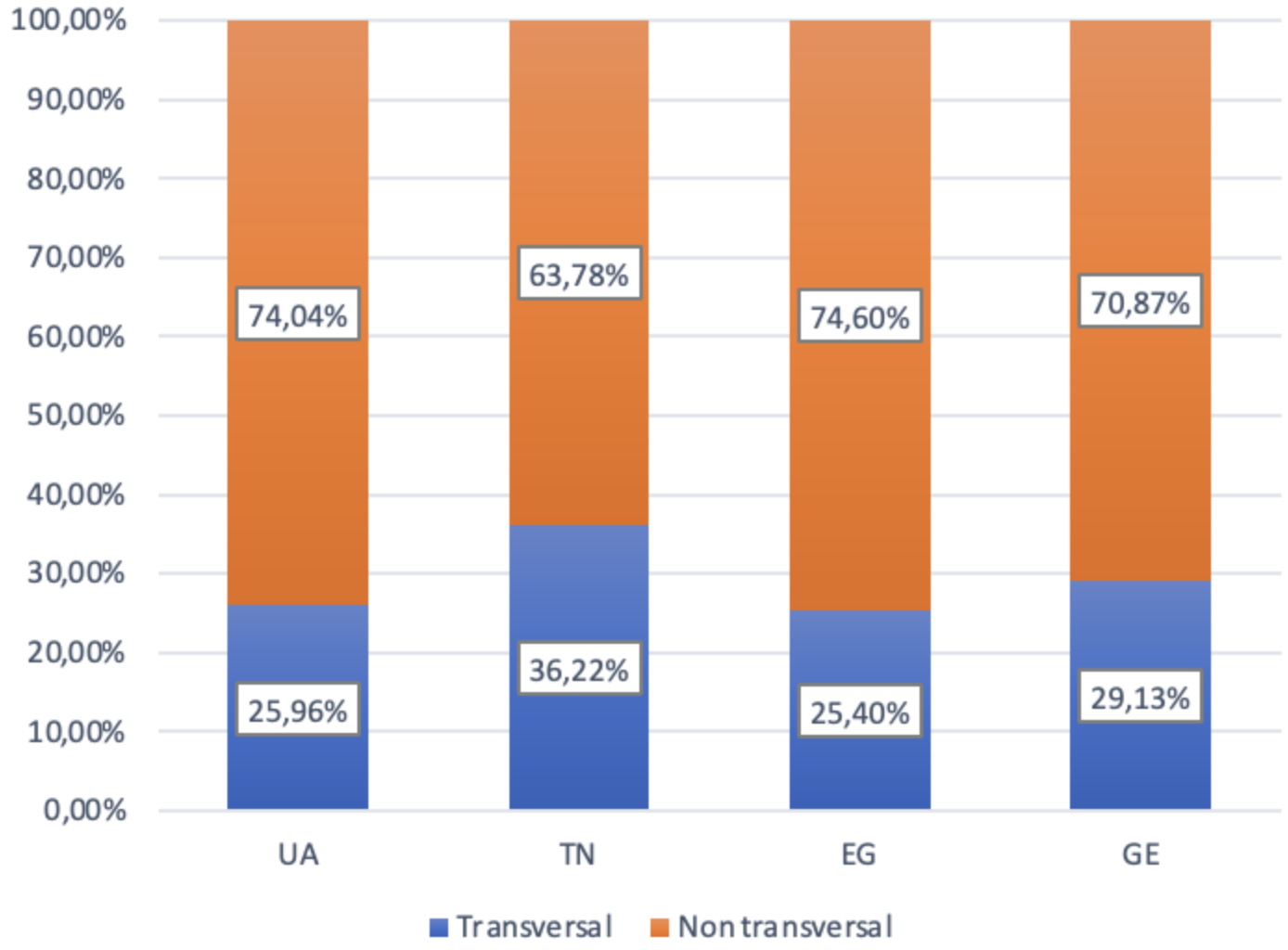
- **Професійно-технічні уміння:** уміння, пов’язані з конкретною посадою, та компетентності, необхідні для виконання певної роботи або завдання (наприклад, володіння певним ПЗ або інструментами, певні навички ручної роботи і т.д.)
- **Комунікативні уміння:** є наскрізними по своїй суті і стосуються здатності людини взаємодіяти з іншими людьми та з середовищем (наприклад, комунікативні навички, вирішення проблем, і т.д.)

Наскрізні уміння в ESCO

- **Уміння думати.** Здатність використовувати ментальні процеси та мислення для вирішення складних проблем, збільшення знань та виконання складних завдань.
- **Соціальна взаємодія.** Здатність розвивати стосунки, взаємодіяти та брати участь у спільній діяльності з колегами, клієнтами та споживачами.
- **Застосування знань.** Загальне застосування умінь, що зазвичай використовуються на роботі та в навчанні; розуміння організації та робочого середовища.
- **Ставлення та цінності.** Стиль роботи людини, її уподобання та переконання, пов'язані з роботою, які лежать в основі поведінки. Наприклад, здатність до змін, до самостійної роботи, здатність дотримуватись взятих на себе зобов’язань.

Ми включаємо у наскрізні вміння також знання мов.

Уміння: порівняння наскрізних та не наскрізних умінь



ETF Working together Learning for life
European Training Foundation

Найважливіші наскрізні уміння - Україна

Найважливіші уміння, що не є наскрізними - Україна

Фактори ринку праці, що є викликом

Lightcast

Ринок праці - фактори, що є викликом

- Еволюція умінь
- Нові професії, що виникають
- Автоматизація роботи/заміна
- …..

Розрахунок коефіцієнтів умінь та показників актуальності умінь

Коефіцієнт уміння відображає частоту цифрових компетентностей, нецифрових професійно-технічних умінь та комунікативних умінь, що поєднуються в одній професії.

Показник актуальності уміння відображає вагу кожного уміння в даній професії

Ідея: Використати інформативність ОДВ для розрахунку коефіцієнтів уміння та показників актуальності уміння

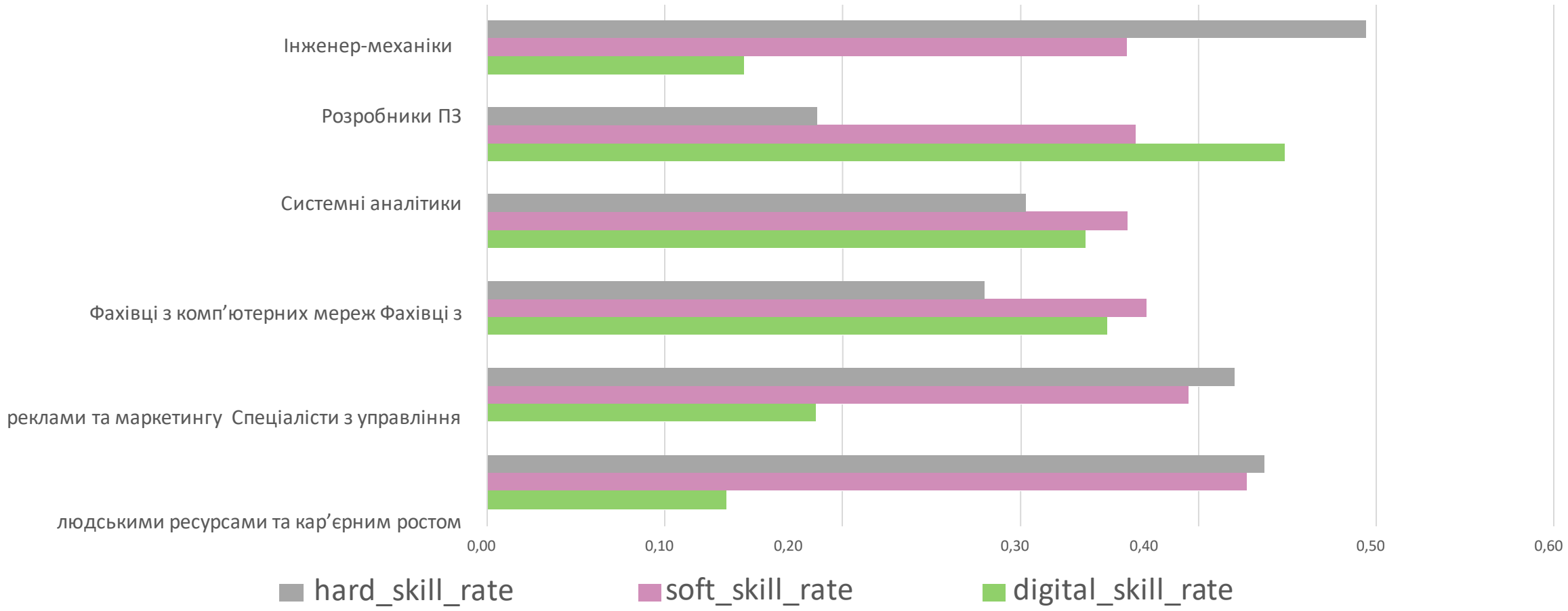
Методологія коефіцієнтів умінь та показників актуальності була розроблена та використана CRISP у 2017 р. та 2018 р. для Італійської обсерваторії цифрових умінь за підтримки італійських асоціацій ІКТ.

Визначення та метод розрахунку

$$\text{Коефіцієнт} = \frac{(\ h\ t)}{+\ h\ +}$$

$$= \frac{^{\circ}\ (,l)}{^{\circ}\ ()} \times \frac{^{\circ}\ (,\mathrm{l})}{^{\circ}\ (\mathrm{n})}$$

Коефіцієнти умінь для деяких професій



Заключні зауваження

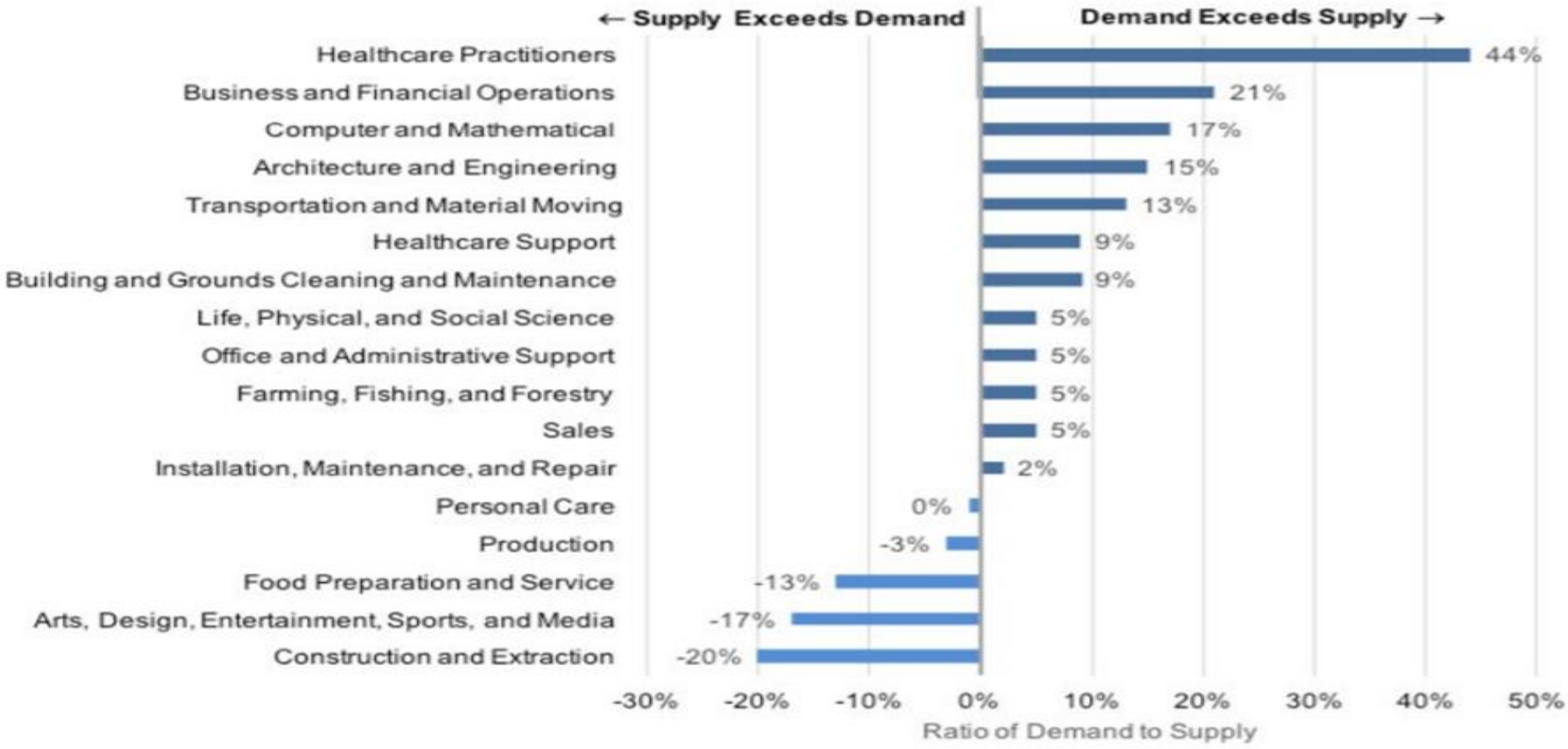
- Знання є критично важливим при відслідковуванні факторів ринку праці, що є викликом:
  - ✓ Еволюція умінь
  - ✓ Нові професії, що виникають
  - ✓ Автоматизація роботи/заміна
  - ✓ …
- ОДВ доповнюють інші ресурси, які ми звичайно використовуємо для того, щоб зрозуміти те, що відбувається на ринку праці
- Ми лише починаємо використовувати інформаційний потенціал ОДВ та великих даних

Розуміння великих даних

Lightcast

Не завжди недостатня кількість означає дефіцит

Поглиблена характеристика дефіциту вмінь



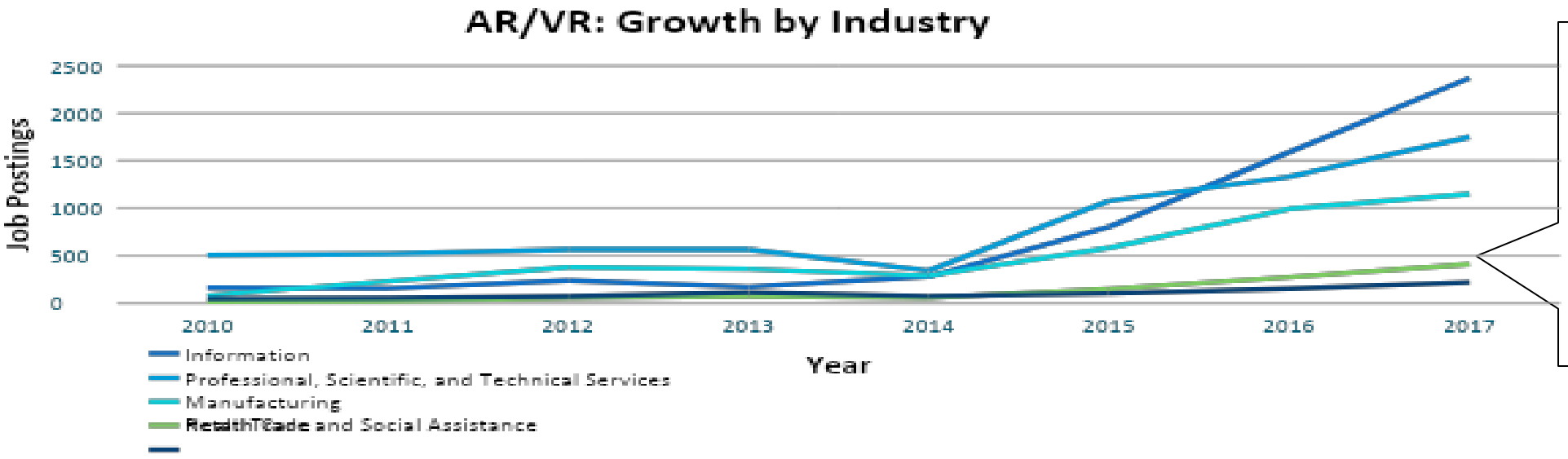
European Training Foundation

Визначення потреб в уміннях у майбутньому

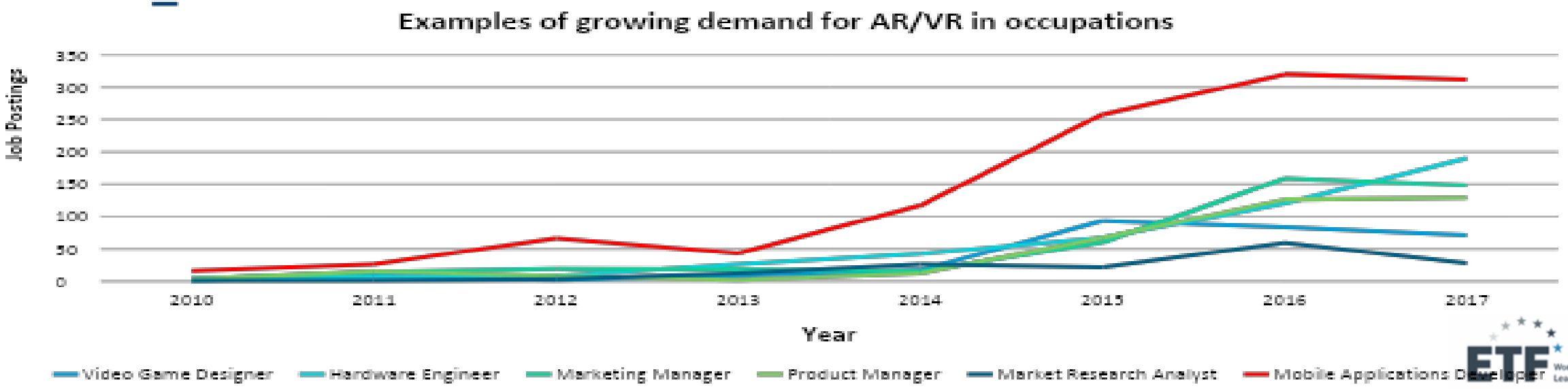
Набір інструментів для відслідковування нових тенденцій

	Головні уміння в галузі IT (Загальна кількість оголошень)	Найбільш оплачувані вміння в галузі IT (Середня зарплатня, вказана в оголошеннях)	Уміння в галузі IT, що зростають найшвидше (Прогноз на 24 місяці)	Уміння в галузі IT, які знайти найважче (Середня тривалість оголошення)
1	SQL	Zookeeper	TensorFlow	Безпека публічних хмарних сервісів
2	Java	TensorFlow	Загальний регламент про захист даних (GDPR)	Інфраструктура як послуга (IaaS)
3	JavaScript	Scala	Kubernetes	Архітектура хмарної обробки даних
4	Linux	AWS Redshift	Spring Boot	Хмарна інфрастуктура
5	Python	AWS DynamoDB	Webpack	Ansible
6.	Аналітика даних	Go Programming Language (Golang)	AWS Lambda	Apache Mesos
7.	Сфера продажу	Pig	Salesforce Lightning	Data Protection Planning
8.	C#	Apache Mesos	Redux	Структура розподілу роботи
9.	Scrum	AWS CloudFormation	Financial Microservices	Hadoop Cloudera
10.	C++	Deep Learning	Apache Kafka	OpenShift

Оцінка потреб у вміннях у нових галузях



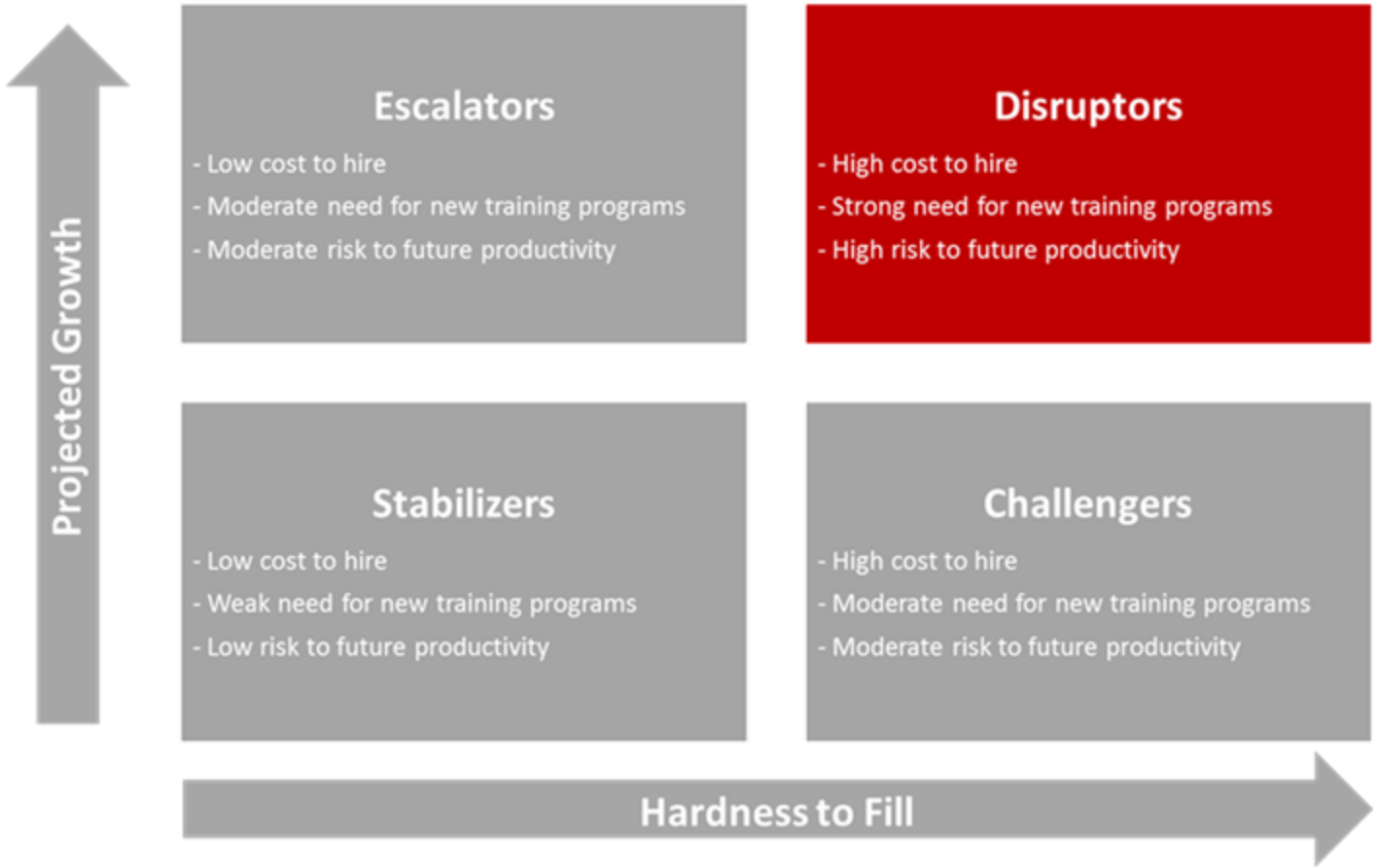
В галузі роздрібної торгівлі, до якої значною мірою належать споживачі AR/VR (технологій доповненої/віртуальної реальності), з 2015 р. попит зростає пришвидшеними темпами.



ETF Working together Learning for life
European Training Foundation

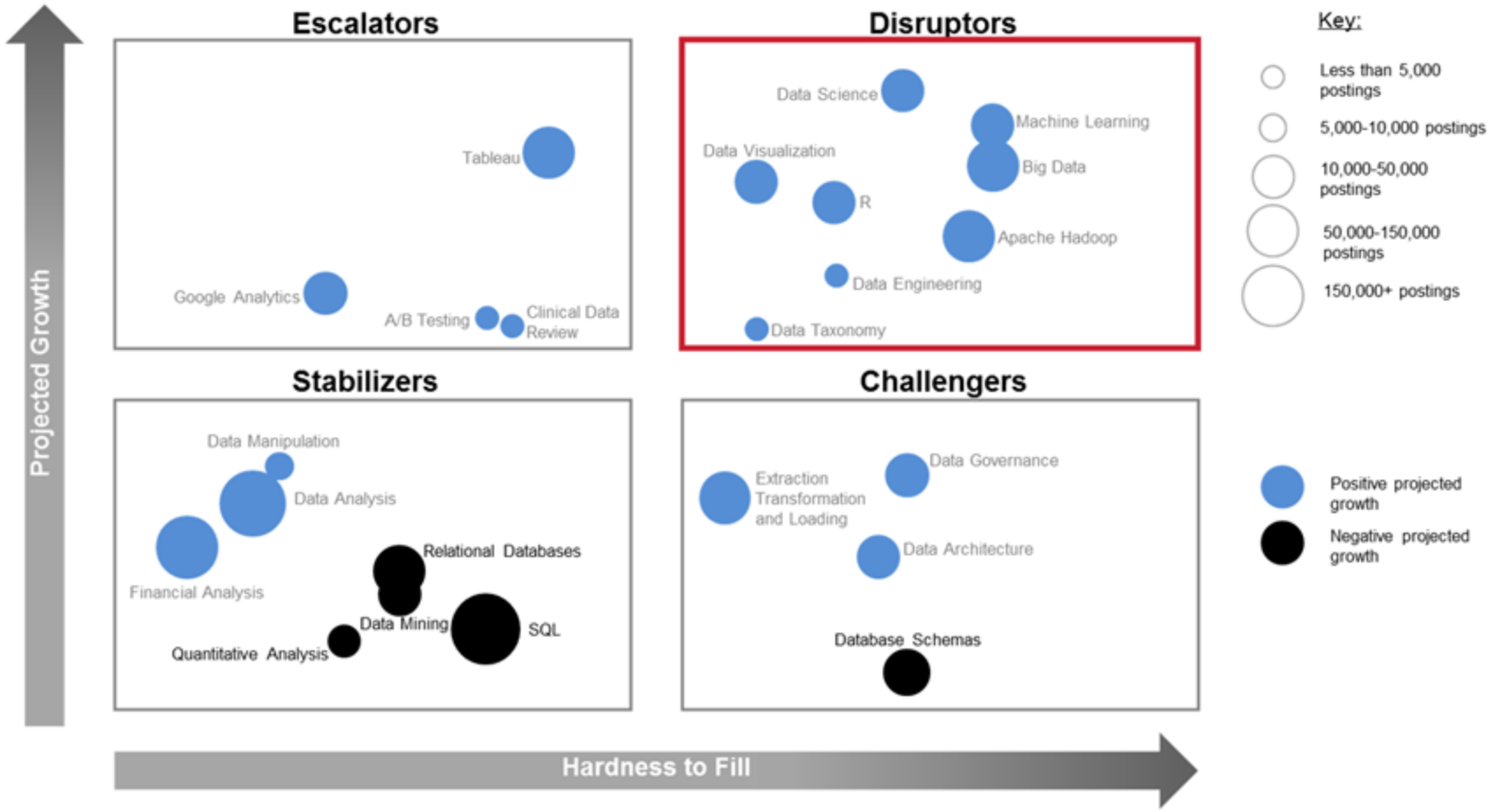
Визначення та пріоритизація видів роботи та умінь проривного характеру

Структура для визначення та завчасного планування умінь, які вірогідно стануть викликом для ринку праці у майбутньому.



ETF Working together Learning for life
European Training Foundation

Навіть, у таких нових галузях, як аналіз даних, не всі уміння, що створюються, мають однакове значення

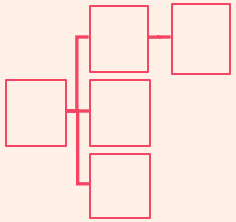


ETF Working together Learning for life
European Training Foundation

Поєднання різних джерел

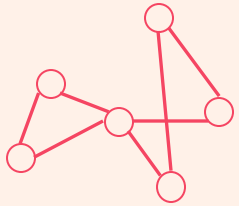
Lightcast

Структуровані дані у порівнянні з великими даними



Структуровані дані

Спеціально зібрані та співставлені дані, що містяться у чіткій впорядкованій структурі. У галузі виробництва електронних виробів це, як правило, дані з державних статистичних оглядів, які містять відповіді на чіткі запитання, поставлені до цільової вибірки конкретної аудиторії.



Великі дані

Надзвичайно великий масив даних, зібраних на основі різних операцій, а не в результаті цілеспрямованої діяльності зі збору даних. У випадку галузі виробництва електронних виробів це означає збір оголошень про роботу та профілів працівників з різних інтернет-ресурсів.

Зведені дані у порівнянні з мікроданими

Зведені дані

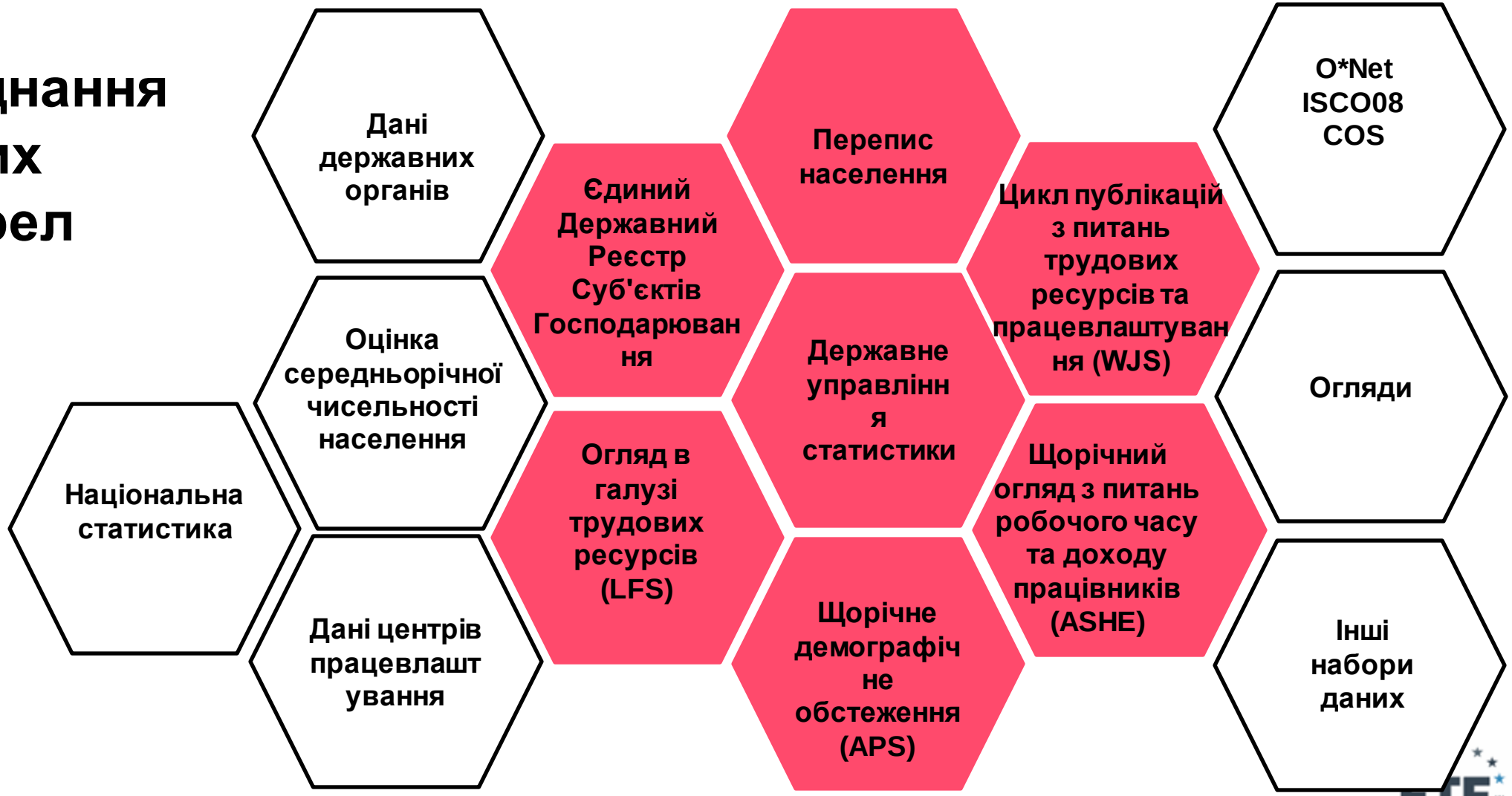
Зведені дані ми отримуємо у вигляді узагальненої статистики: загальна кількість X, середній рівень Y, відсоток Z. Багато які структуровані дані доступні лише у зведеній формі з обмеженим доступом до мікроданих.

Мікродані

Мікродані дають більші можливості: це дані, які наявні для кожного окремого випадку без узагальнення, і ми можемо потім обробляти їх як нам потрібно. Великі дані поступають до нас у формі мікроданих, хоча часто ми здійснюємо їх підготовку для стандартних форм зведення.

Lightcast

Поєднання різних джерел



ETF Working together Learning for life
European Training Foundation

Lightcast

Аналітика оголошень про роботу та професійних профілів - аналіз ринку праці на основі документів

- **Аналітика оголошень про роботу та професійних профілів є аналізом ринку на основі документів**: це джерело великих даних, де весь процес перебуває під нашим контролем, включаючи необроблені дані.

- **Аналітика оголошень про роботу та професійних профілів відбувається (майже) в режимі реального часу**: ми виконуємо скрейпінг оголошень про роботу кожен день, і вже за день чи два ми можемо використовувати ці дані для аналізу.

- **Аналітика оголошень про роботу та професійних профілів має широкі можливості**: за останні 2 роки було опубліковано мільйони оголошень про роботу, і ця кількість зростає.

- **Аналітика оголошень про роботу та професійних профілів містить шум**: оголошення про роботу публікуються по різному в залежності від попиту на ринку праці і це потрібно враховувати.

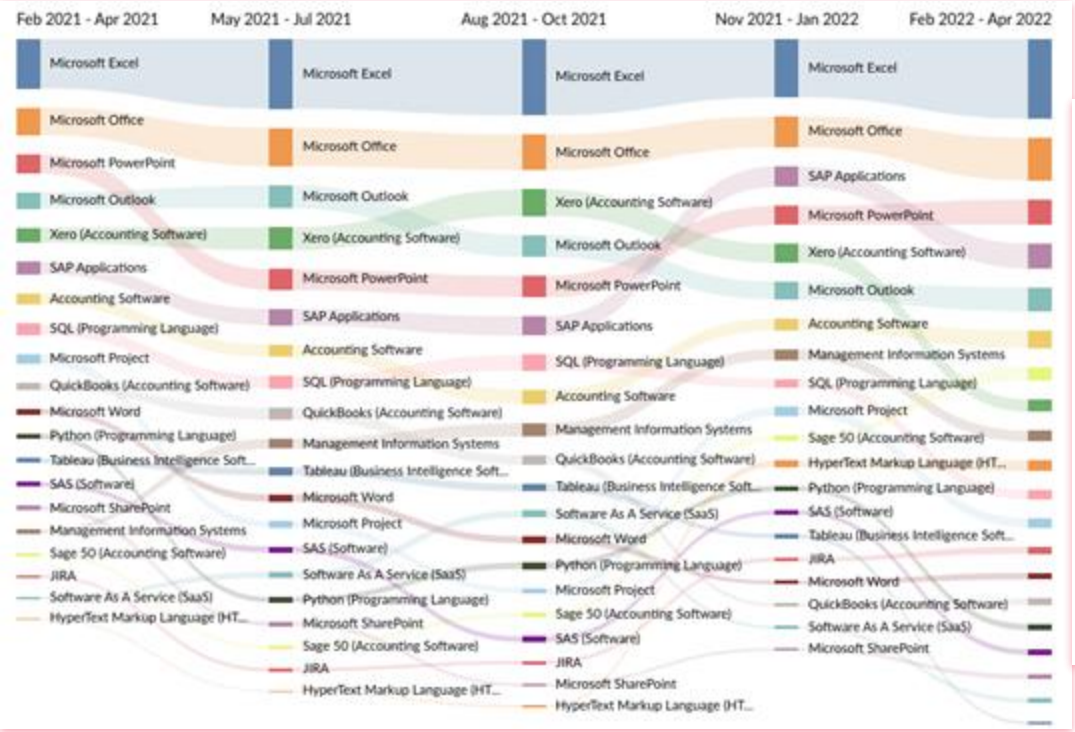
ETF Working together Learning for life
European Training Foundation

Інтерактивне заняття

ETF Working together Learning for life
European Training Foundation

Тенденції в сфері умінь | Дефіцит умінь | Діяльність з найму
від Lightcast Analyst (Сполучене Королівство)

Професії, віднесені до сфери курсу «Бухгалтерський облік» (JACS N400) у містах Брістоль, Бат і Північно-Східний Сомерсет, Північний Сомерсет та Південний Глостершир, квітень 2021 р. – травень 2022 р.



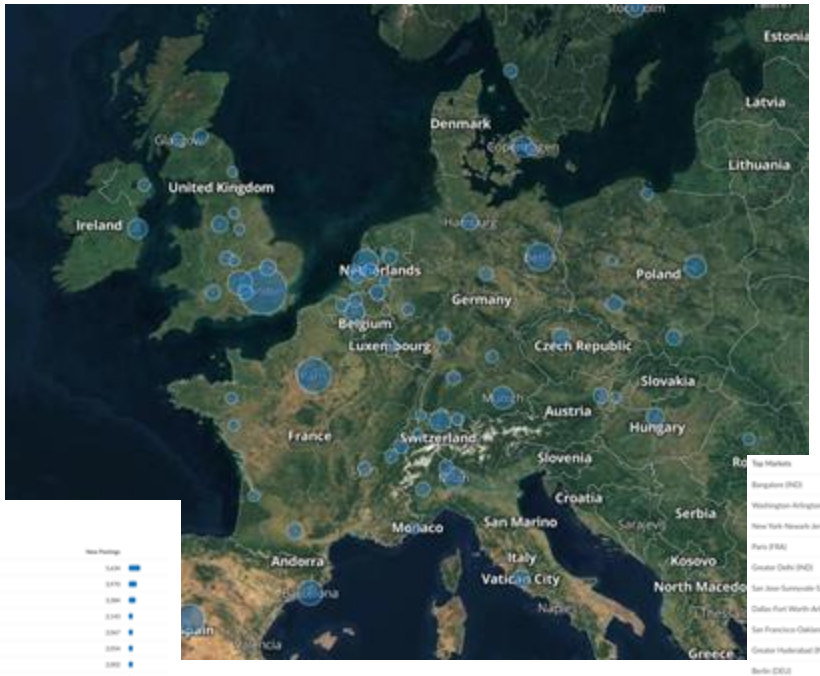
ETF Working together Learning for life
European Training Foundation

Розуміння вмінь на міжнародному рівні

від Lightcast Global

Наприклад, аналітики та фахівці по роботі з даними з умінням програмувати на Python

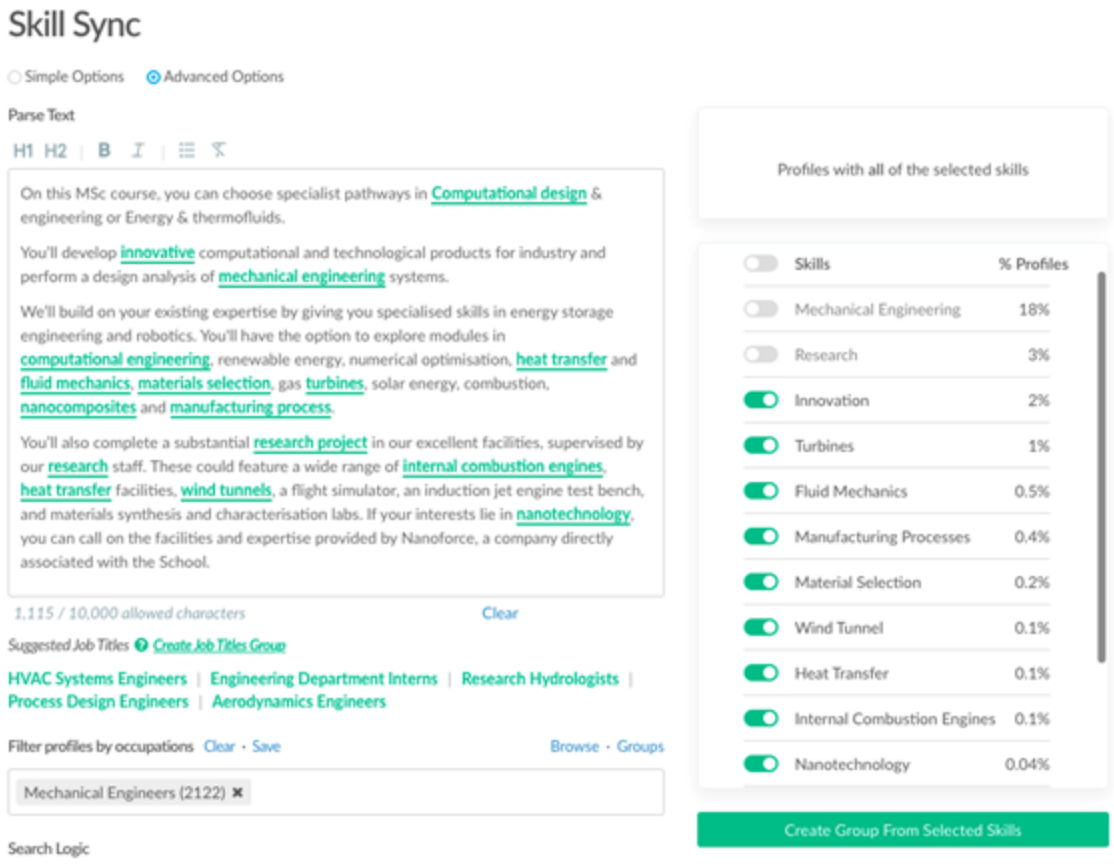
Визначити сфери, в яких зосереджені пропозиція та попит на професії в усьому світі, і поглиблено проаналізувати дані для того, щоб набути розуміння того, якими є найважливіші уміння та роботодавці для певної групи населення та зробити оцінку цієї групи населення.

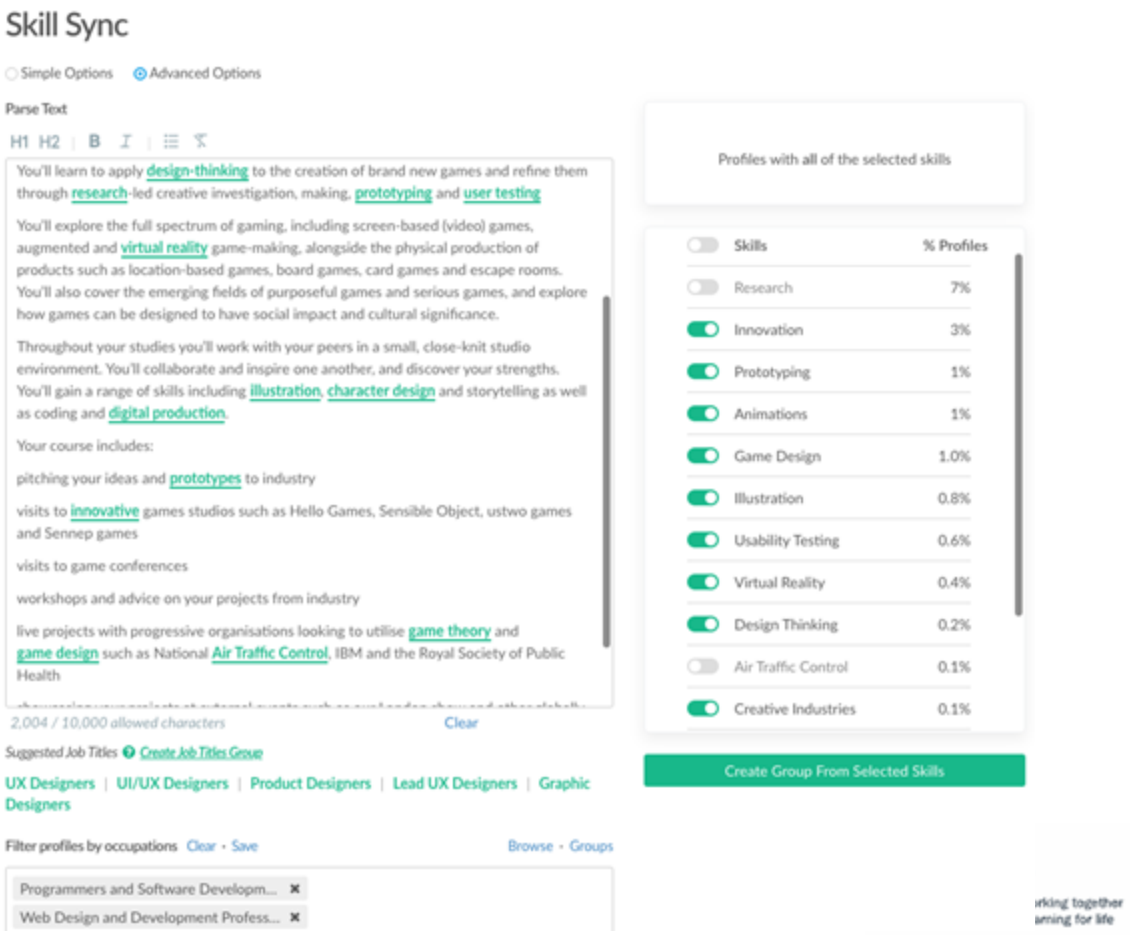


Перевірити дані стосовно вмінь в документації курсу

за допомогою Lightcast Skill Sync

Наприклад, магістерська програма «Сучасне машинобудування» (коледж Куін Мері Лондонського університету), бакалаврська програма «Дизайн та мистецтво ігор»





Start Over

Timeframe

Aug 2018 to Sep 2022

1 Occupation Advanced · Clear · Save

Results should...

Include

× Programmers and Software Developmen...

Add Occupations...

1 Nation Advanced · Clear · Save

× United Kingdom (K02000001)

Add Regions...

Job Titles

None selected.

Add Job Titles...

Companies Groups

None selected.

Add Companies...

Include Skills & Qualifications

Has ANY of the following skills

Groups

None selected.

Add Skills & Qualifications...

- AND -

Has ALL of the following skills

Groups

None selected.

Add Skills & Qualifications...

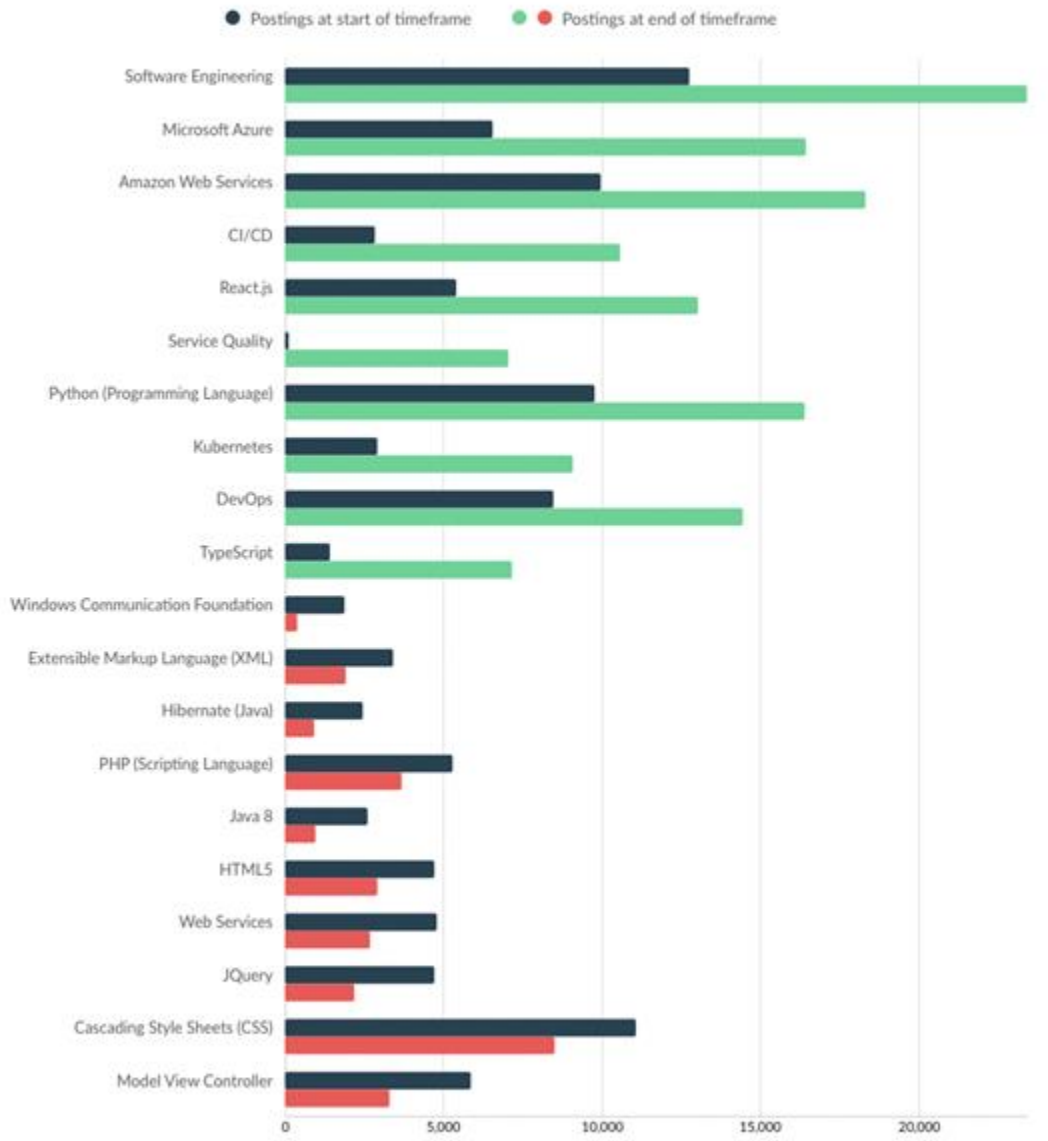
Exclude Skills & Qualifications Groups

Results should...

Exclude any of the selected skills

Hot and Cold Skills by Job Postings

Save Export



Skill	Postings in Aug 2018	Postings in Sep 2022	Change	Percent Change

ETF Working together Learning for life

European Training Foundation