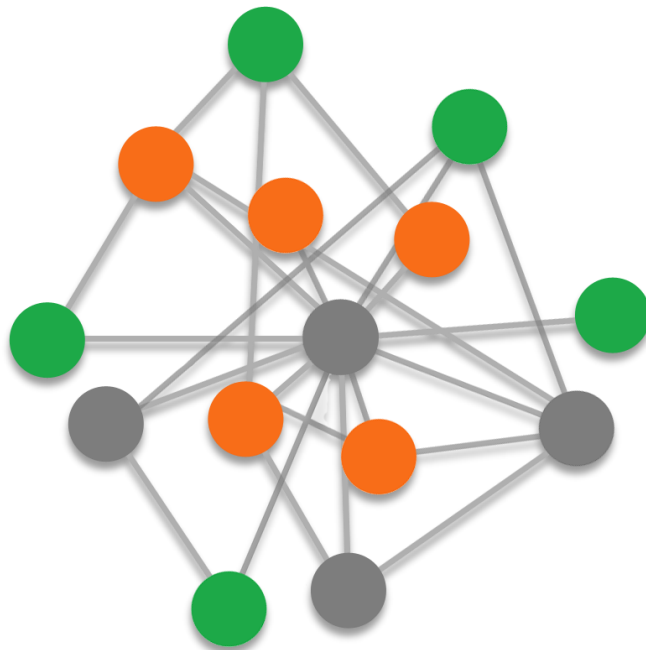


BELHAOUZIA Zohir
LAUNAY Swan
GENTER Willy
JUBARD Théo

INFO906
Vendredi 9 Décembre 2022

INFO906 : Web Sémantique



2022-2023

Table des matières

- 1- Définition des objectifs du projet et des questions de compétences³**
- 2- Identification du ou des graphes de connaissances⁴**
 - a- Choix du graphe est domaine couvert⁴**
 - b- Vocabulaire utilisés⁴**
- 3- Extraction d'informations à partir du graphe de connaissances⁶**
 - a- La plateforme Wikidata⁶**
 - Wikidata Wiki⁶
 - Wikidata Query Builder⁶
 - Wikidata Query Service⁶
 - b- Les requêtes SPARQL⁷**
 - Quels sont les genres / sexes des Jedi ?⁷
 - Quels sont les relations maître / padawan entre les Jedi ?⁸
 - Par quels acteurs sont interprétés les Jedi et quelles sont les notations des films ?⁹
- 4- Traitement(s) des résultats retournés par les requêtes SPARQL¹¹**
 - a- Processus de représentation et d'analyse des données¹¹**
 - b- Résultats par questions¹²**
 - Quels sont les genres / sexes des jedis ?¹²
 - Par quels acteurs sont interprétés les Jedi et quelles sont les notations ?¹³
 - Quels sont les relations maître / padawan entre les Jedi ?¹⁵

1- Définition des objectifs du projet et des questions de compétences

Dans le cadre du module de Web Sémantique et de l'étude des ontologies et ainsi que des graphes de connaissances, nous avons décidé de nous pencher sur le cas de l'univers cinématographique de Star Wars et plus précisément les Jedi.

Ce n'est pas seulement l'univers en lui-même qui sera étudié mais aussi l'impact de cette saga sur les différents individus ayant participé à son rayonnement comme les acteurs présent dans les films.

Afin de faire ressortir les éléments tels que les relations, propriétés et classes présente dans le graphe de connaissances étudiés, plusieurs questions de compétences ont été formulées.

Pour rappel, une question de compétence peut être définie de la manière suivante :

« Une question de compétence est une phrase du langage naturel exprimant un motif de questionnement pour lequel il est attendu qu'une ontologie apporte une réponse »

C'est avec cette définition en tête ainsi que le sujet de notre projet que les 3 questions de compétences suivantes ont été formulées :

- Quels sont les genres / sexes des Jedi ?
- Par quels acteurs sont interprétés les Jedi et quelles sont les notations des films ?
- Quels sont les relations maître / padawan entre les Jedi ?

Afin de répondre à ces questions, 3 requêtes Sparql ont été formulées afin d'être utilisées sur un graphe de compétence adéquate. Le reste de ce compte-rendu traitera le choix du graphe de connaissances, de la forme des requêtes Sparql ainsi que des résultats obtenus, leur représentation et analyse ainsi que des outils et méthodes utilisés pour représenter/obtenir ces résultats.



2- Identification du ou des graphes de connaissances

a- Choix du graphe est domaine couvert

Afin d'avoir le plus d'informations possibles, notre recherche de graphe s'est portée sur ceux traitant de l'univers étendu de Star Wars, c'est-à-dire les films mais aussi les comics, bande dessinée ou encore jeu vidéo. En dehors de l'univers en lui-même, des informations relatives à la production de ces œuvres et à la notation de ces œuvres étaient aussi nécessaire afin de répondre à notre question sur la relation entre l'univers Star Wars et les acteurs ayant prêté leurs traits ou voix à des personnages de la saga.

C'est avec ces éléments en tête que le graphe de connaissance [Wikidata](#) a été choisie.

Graphe de connaissance utilisé par Wikipédia, Wikidata est une base de connaissance libre éditée et enrichie de manière collaborative par une large communauté. Servant de base à de nombreux articles de Wikipédia, cette base a le mérite d'être imposante et de répondre à nos attentes en termes de contenu tout en proposant celui-ci dans différentes langues. De plus, Wikidata vient avec un certain nombre d'outils permettant la visualisation de données, la recherche de celle-ci ainsi que son extraction de la base.

b- Vocabulaire utilisés

Comme beaucoup des graphes présentés dans ce module, il est possible d'effectuer des requêtes SPARQL via un endpoint dédié sur Wikidata.

Une grande partie du vocabulaire ainsi présenté est présent dans les systèmes de requêtes comme **rdf**, **skos** ou **owl**.

En-dehors de ces vocabulaires déjà connus, Wikidata embarque son propre vocabulaire permettant ainsi de représenter et de rechercher les entités ainsi que leurs relations et propriétés. Tous les éléments du vocabulaire n'ont pas été utilisés dans le cadre de ce projet mais les suivants peuvent être retrouvés dans nos requêtes SPARQL :

wdt (référence à une relation), **wd** (référence à l'entité qui doit directement être liée à la relation), **p** (référence à une propriété d'une entité), **ps** (référence à la valeur de la propriété), **bd** (permet de choisir le graphe qui sera utilisé ainsi que la langue de recherche des informations).

En dehors de ces éléments de vocabulaire, d'autres préfixes viennent enrichir les options mises à disposition afin d'effectuer des recherches plus précises sur les valeurs de propriétés ou encore sur les relations qui les relient à leurs entités. Voici la liste du vocabulaire spécifiques à Wikidata ainsi que des préfixes non-spécifiques utilisables avec cette base :

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wds: <http://www.wikidata.org/entity/statement/>
PREFIX wdv: <http://www.wikidata.org/value/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX wikibase: <http://wikiba.se/ontology#>
PREFIX p: <http://www.wikidata.org/prop/>
PREFIX ps: <http://www.wikidata.org/prop/statement/>
PREFIX pq: <http://www.wikidata.org/prop/qualifier/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX bd: <http://www.bigdata.com/rdf#>
```

```
PREFIX wdref: <http://www.wikidata.org/reference/>
PREFIX psv: <http://www.wikidata.org/prop/statement/value/>
PREFIX psn: <http://www.wikidata.org/prop/statement/value-normalized/>
PREFIX pqv: <http://www.wikidata.org/prop/qualifier/value/>
PREFIX pqn: <http://www.wikidata.org/prop/qualifier/value-normalized/>
PREFIX pr: <http://www.wikidata.org/prop/reference/>
PREFIX prv: <http://www.wikidata.org/prop/reference/value/>
PREFIX prn: <http://www.wikidata.org/prop/reference/value-normalized/>
PREFIX wdn: <http://www.wikidata.org/prop/novalue/>
PREFIX wdata:
<http://www.wikidata.org/wiki/Special:EntityData/>
```

```
PREFIX schema: <http://schema.org/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX bds: <http://www.bigdata.com/rdf/search#>
PREFIX gas: <http://www.bigdata.com/rdf/gas#>
PREFIX hint: <http://www.bigdata.com/queryHints#>
```

3- Extraction d'informations à partir du graphe de connaissances

a- La plateforme Wikidata

Le site **wikidata.org** met à disposition plusieurs outils pour interroger la base de connaissance Wikidata. L'Endpoint SPARQL de la plateforme permet à n'importe quelle personne de composer des requêtes pour parcourir le graph de connaissance et récupérer les données voulues.

Dans le cadre de ce projet nous avons utilisé 3 outils proposés par la plateforme Wikidata pour visualiser et extraire des données sur le thème des Jedi de l'univers Star Wars.

Wikidata Wiki

Le wiki permet d'explorer le graph de connaissance de manière graphique, c'est-à-dire de la même façon qu'on pourrait naviguer sur internet. Cette navigation intuitive permet d'avoir un aperçu rapide des données que l'on voudra extraire par la suite, cela permet notamment de consulter la liste des relations disponibles et la diversité des propriétés liées à un nœud (entité).

Wikidata Query Builder

Une fois le premier débroussaillage effectué avec le wiki, il faut définir la liste des relations que l'on souhaite utiliser pour parcourir notre graphe. Le Query Builder permet justement d'exécuter des requêtes simples vers la base de connaissance en proposant un système d'autocomplétions des concepts et propriétés. Cette exécution rapide de requêtes permet de tester de façon indépendante les différentes instructions qui composeront la requête finale.

Wikidata Query Service

Une fois notre requête prête, elle peut être exécutée dans le Query Service qui va interpréter notre code et nous renvoyer les données associées issus de la base de connaissance Wikidata. Le Query Service permet également de visualiser ou de convertir les données ainsi récupérées pour :

1. Constater que la requête s'est bien déroulée et que le résultat est satisfaisant
2. Pouvoir exporter les données dans des formats connus pour faciliter leur exploitation avec d'autres langages de programmation/outils en ligne.

b- Les requêtes SPARQL

Dans cette section, les requêtes seront présentées et décrites par question de compétences.

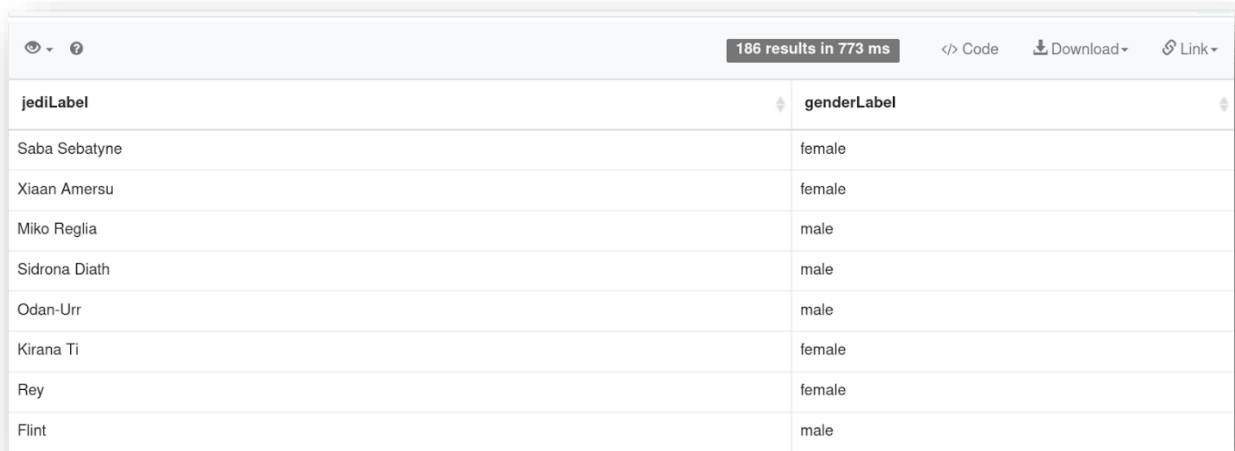
Quels sont les genres / sexes des Jedi ?

La requête :

```
SELECT DISTINCT ?jediLabel ?genderLabel WHERE {  
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }  
  {  
    SELECT ?jedi ?gender WHERE {  
  
      {?jedi p:P106 ?isJedi0.  
       ?isJedi0 (ps:P106/(wdt:P279*)) wd:Q51724.}  
      UNION  
      {?jedi p:P463 ?isJedi1.  
       ?isJedi1 (ps:P463/(wdt:P279*)) wd:Q51724.}  
  
      ?jedi wdt:P21 ?gender.  
    }  
  }  
}
```

Commençons par un exemple simple. La requête ci-dessus permet de récupérer la liste de tous les Jedi avec leur genre. On commence par récupérer toutes les entités considérées comme des Jedi. Elles sont identifiées tel quel si elles sont membres (**member_of (P106)**) ou si elles occupent un poste (**occupation (P463)**) à l'ordre des Jedi (**Q51724**).

Une fois l'union de ces 2 ensembles effectués, on obtient le résultat suivant (la capture ci-dessous est un fragment du résultat) :



jediLabel	genderLabel
Saba Sebatyne	female
Xiaan Amersu	female
Miko Reglia	male
Sidrona Diath	male
Odan-Urr	male
Kirana Ti	female
Rey	female
Flint	male

Quels sont les relations maître / padawan entre les Jedi ?

La requête :

```
SELECT DISTINCT ?jediLabel ?studentLabel WHERE {
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
  {
    SELECT ?jedi ?student WHERE {
      {?jedi p:P106 ?isJedi0.
        ?isJedi0 (ps:P106/(wdt:P279*)) wd:Q51724.}
      UNION
      {?jedi p:P463 ?isJedi1.
        ?isJedi1 (ps:P463/(wdt:P279*)) wd:Q51724.}
    }
    ?jedi wdt:P802 ?student.
  }
}
```

Pour notre deuxième requête, on cherche à récupérer les élèves (padawans) de nos Jedi.

La requête commence de la même façon que la précédente pour récupérer les Jedi, puis on utilise la relation **Student (P802)** pour récupérer les élèves de chaque Jedi. Il faut noter que comme la clause de récupération des élèves n'est pas optionnelle, la requête omettra les éventuels Jedi qui n'ont pas de padawan. De plus, il est tout à fait possible que certains individus aient eu le rôle de padawan et soit maintenant maître.

Une fois la requête exécutée, on obtient les résultats suivants :

131 results in 982 ms		Code	Download	Link
jediLabel	studentLabel			
Luke Skywalker	Kyp Durron			
Luke Skywalker	Tionne Solusar			
Leia Organa	Allana Solo			
Exar Kun	Ulic Qel-Droma			
Darth Traya	Darth Sion			
Yoda	Ki-Adi-Mundi			
Yoda	Cin Drallig			

Par quels acteurs sont interprétés les Jedi et quelles sont les notations des films ?

La requête :

```
SELECT DISTINCT ?jediLabel ?actorLabel ?filmLabel ?scoreLabel WHERE
{
  SERVICE wikibase:label { bd:serviceParam wikibase:language
"[AUTO_LANGUAGE]". }
  {
    SELECT ?jedi ?actor ?film ?score WHERE {

      #Get Jedis (member_of or occupation of Jedi)
      {?jedi p:P106 ?isJedi0.
      ?isJedi0 (ps:P106/(wdt:P279*)) wd:Q51724.}
      UNION
      {?jedi p:P463 ?isJedi1.
      ?isJedi1 (ps:P463/(wdt:P279*)) wd:Q51724.}

      #Get actors linked to Jedis
      {?jedi p:P175 ?isActor.
      ?isActor (ps:P175/(wdt:P279*)) ?actor.}

      #Get films linked to actor...
      {?film p:P161 ?isFilm.
      ?isFilm (ps:P161/(wdt:P279*)) ?actor.}

      #... that are in the StarWars fictional universe
      {?film p:P1434 ?isInFiction.
      ?isInFiction (ps:P1434/(wdt:P279*)) wd:Q19786052.}

      #Get the film rating if exists
      OPTIONAL
      {?film wdt:P444 ?score.}
    }
  }
}
```

La requête ci-dessus permet de lier les Jedi (des personnages fictifs) aux acteurs qui les ont interprétés, puis aux films (de l'univers Star Wars) dans lesquels ces derniers ont joué un rôle. L'idée derrière la récupération de ces informations est d'analyser l'impact que la participation en tant que Jedi à la saga Star Wars a pu avoir sur la carrière des acteurs.

Décomposons un peu la requête :

Récupération des Jedi :

- Idem que pour les requêtes précédentes

Liaison avec les acteurs :

- On utilise la relation **Performer (P175)** liée à un Jedi pour accéder à son interprète.

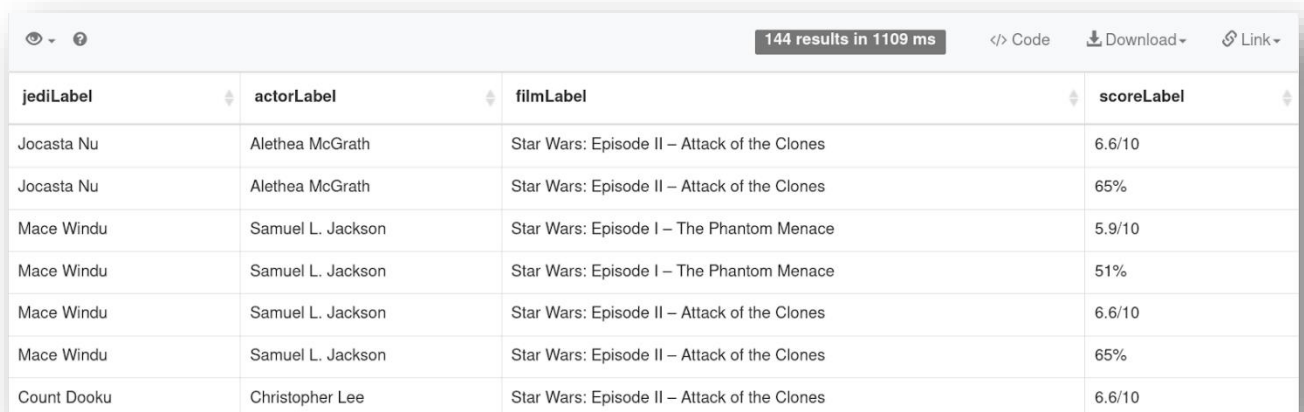
Les films de la saga Star Wars auxquels les acteurs ont participé :

- On récupère tous les films appartenant à la saga Star Wars, c'est-à-dire ayant une relation **Takes Place in Fictional Universe (P1434)** avec l'entité **StarWars Universe (Q19786052)**.
- Puis on fait le lien entre les acteurs précédemment récupérés et les films de la Saga via la relation **Cast Member (P175)**

Notation des films (si elle existe) :

- Pour terminer, on récupère la notation de chaque film avec la relation **Review Score (P444)** ce qui nous permet d'associer un film à une note (en pourcentage ou ramenée sur 10 en fonction des cas)

L'exécution complète de la requête donne le résultat suivant :



The screenshot shows a query result interface with a table of 4 columns: jediLabel, actorLabel, filmLabel, and scoreLabel. The table contains 7 rows of data. Above the table, it indicates '144 results in 1109 ms' and provides options for Code, Download, and Link.

jediLabel	actorLabel	filmLabel	scoreLabel
Jocasta Nu	Alethea McGrath	Star Wars: Episode II – Attack of the Clones	6.6/10
Jocasta Nu	Alethea McGrath	Star Wars: Episode II – Attack of the Clones	65%
Mace Windu	Samuel L. Jackson	Star Wars: Episode I – The Phantom Menace	5.9/10
Mace Windu	Samuel L. Jackson	Star Wars: Episode I – The Phantom Menace	51%
Mace Windu	Samuel L. Jackson	Star Wars: Episode II – Attack of the Clones	6.6/10
Mace Windu	Samuel L. Jackson	Star Wars: Episode II – Attack of the Clones	65%
Count Dooku	Christopher Lee	Star Wars: Episode II – Attack of the Clones	6.6/10

Chaque ligne correspond à un chemin distinct par lequel on pourrait parcourir le graph sans repasser deux fois par le même nœud. L'ensemble des lignes permet de décrire le graph dans sa totalité.

4- Traitement(s) des résultats retournés par les requêtes SPARQL

a- Processus de représentation et d'analyse des données

Afin d'apporter les réponses les plus pertinentes à nos questions, il a été décidé de se baser sur une représentation visuelle des réponses à nos requêtes SPARQL, celle-ci étant des sous-graphes issus de Wikidata.

Pour cela, les requêtes ont été effectuées depuis un script python permettant la transformation des données puis leurs représentations visuelles.

Pour chacune des requêtes, la procédure suivante a été suivie:

- Exécution de la requête Sparql et récupération du résultat sous format JSON à l'aide de la librairie [SPARQLWRAPPER](#).
- Transformation des données en format CSV pour traiter les données.
- Utilisation de la librairie [pyvis](#) pour créer les graphes. Cette librairie nous permet de générer un fichier HTML comportant un graphe.
- Ces fichiers sont situés à la racine de notre dossier Github. Celui-ci permet de servir ces pages directement sur le web et ainsi de rendre accessibles les graphes aux adresses suivantes :
 - Relation Maitre/élève :
<https://mrbellazia.github.io/KnowledgeGrapheJedi/students.html>
 - Répartition des genres parmi les Jedi :
<https://mrbellazia.github.io/KnowledgeGrapheJedi/genders.html>
 - Relation entre les personnages et leurs acteurs :
<https://mrbellazia.github.io/KnowledgeGrapheJedi/actorfilm.html>

En-dehors de c'est page web, le répertoire GitHub contient les résultats au format csv de nos requêtes ainsi que les scripts python ayant été utilisé afin d'obtenir ces données ainsi que leurs représentations.

Le répertoire est disponible à l'adresse suivante :

<https://github.com/mrBELLAZIA/KnowledgeGrapheJedi>

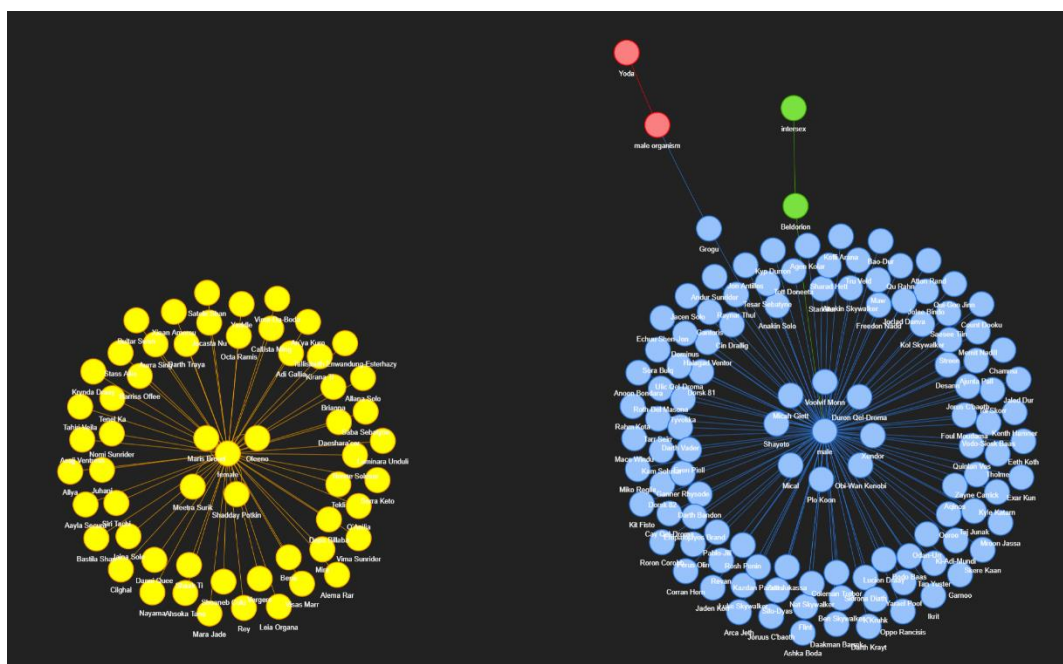
b- Résultats par questions

Quels sont les genres / sexes des jedis ?

Résultat de la requête au format csv

Jedi	Gender
Flint	male
Zayne Carrick	male
Tahiri Veila	female
Rosh Penin	male
Corran Horn	male
Zett Jukassa	male
Tan Yuster	male
Tesar Sebatyne	male
Krynda Draay	female

Représentation sous forme de graphe



Analyse des résultats

Dans un premier temps, voici le code couleur de ce graphe :

En jaune les individus féminins, en bleu les individus masculins, en rouge les organismes masculins et en vert les organismes intersexes.

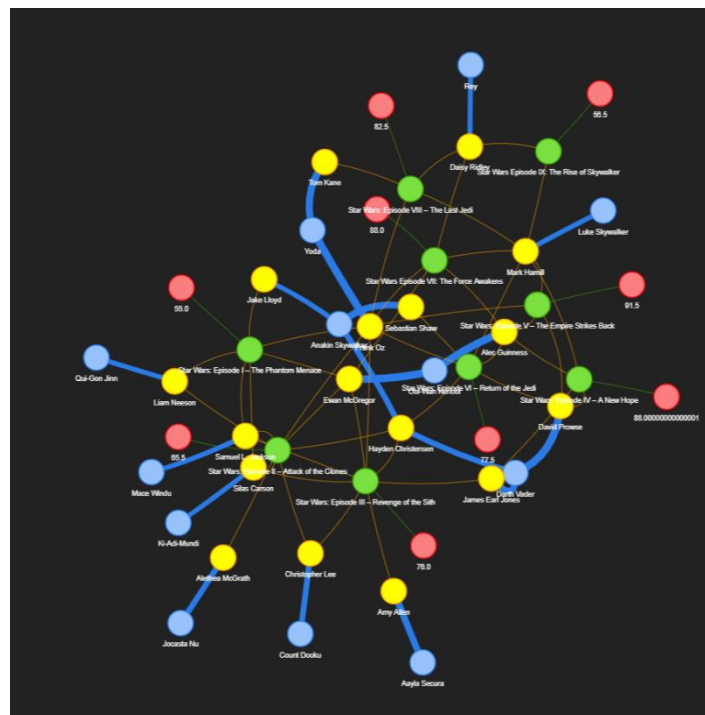
On voit via cette représentation que le nombre d'individus masculins est plus important que le nombre d'individus féminin. De plus, on note la présence de 2 individus considérés comme organisme masculin, cela étant expliqué dans l'univers par la méconnaissance de leur espèce. De plus, le graphe nous indique la présence d'individu intersexe.

Par quels acteurs sont interprétés les Jedi et quelles sont les notations ?

Résultat de la requête au format csv

	jediLabel	actorLabel	filmLabel	scoreLabel
0	Aayla Secura	Amy Allen	Star Wars: Episode III – Revenge of the Sith	76.0
1	Anakin Skywalker	Hayden Christensen	Star Wars: Episode II – Attack of the Clones	65.5
2	Anakin Skywalker	Hayden Christensen	Star Wars: Episode III – Revenge of the Sith	76.0
3	Anakin Skywalker	Hayden Christensen	Star Wars: Episode VI – Return of the Jedi	77.5
4	Anakin Skywalker	Jake Lloyd	Star Wars: Episode I – The Phantom Menace	55.0
5	Anakin Skywalker	Sebastian Shaw	Star Wars: Episode VI – Return of the Jedi	77.5
6	Count Dooku	Christopher Lee	Star Wars: Episode II – Attack of the Clones	65.5
7	Count Dooku	Christopher Lee	Star Wars: Episode III – Revenge of the Sith	76.0

Représentation sous forme de graphe



Analyse des résultats :

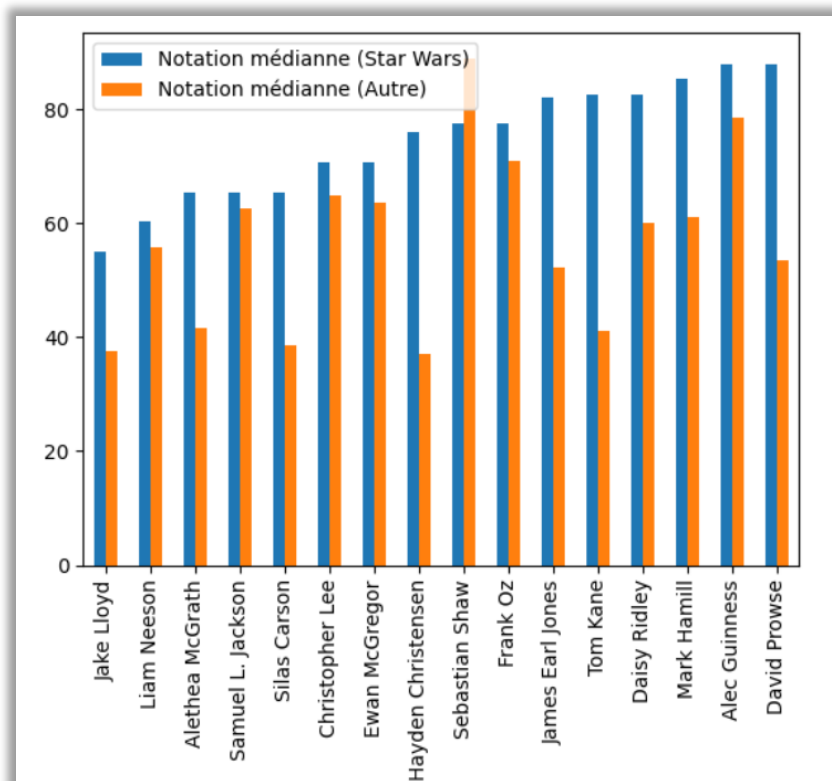
Le code couleur du graphe est le suivant : les entités vertes représentent les films, les entités jaunes les acteurs impliqués dans les films et les entités rouges la notation des films sur 1.

Dans un premier temps, la notation des films nous indique que les épisodes de la prélogie (film I, II et III) sont ceux le moins appréciés de l'univers. Les plus appréciés restent les tout premiers films qui obtiennent les meilleures notes d'après les spectateurs. La dernière trilogie est quant à elle appréciée de manière globale (2 films avec un score de plus de 80%). Pour finir, concernant les notations des films, les premier et derniers épisodes de la saga qui sont le moins appréciés avec les 2 pires notes enregistrées.

Pour ce qui est des acteurs impliqués dans la saga, le graphe nous indique que la plupart de ceux-ci ont joué leurs personnages dans la totalité des films

impliquant leur personnage. Les seules exceptions sont les personnages de Yoda, Darth Vader, Anakin et Yoda et Obi-Wan, ceux-ci étant présents sur la quasi-totalité des films, représentant une période de 42 ans.

Pour finir, une autre partie des informations a permis de définir le score moyen des films dans lesquelles les acteurs présents sont. La représentation graphique suivante montre que la grande majorité de ces acteurs à connu un impact positif de cette Saga dans leurs carrières cinématographiques.

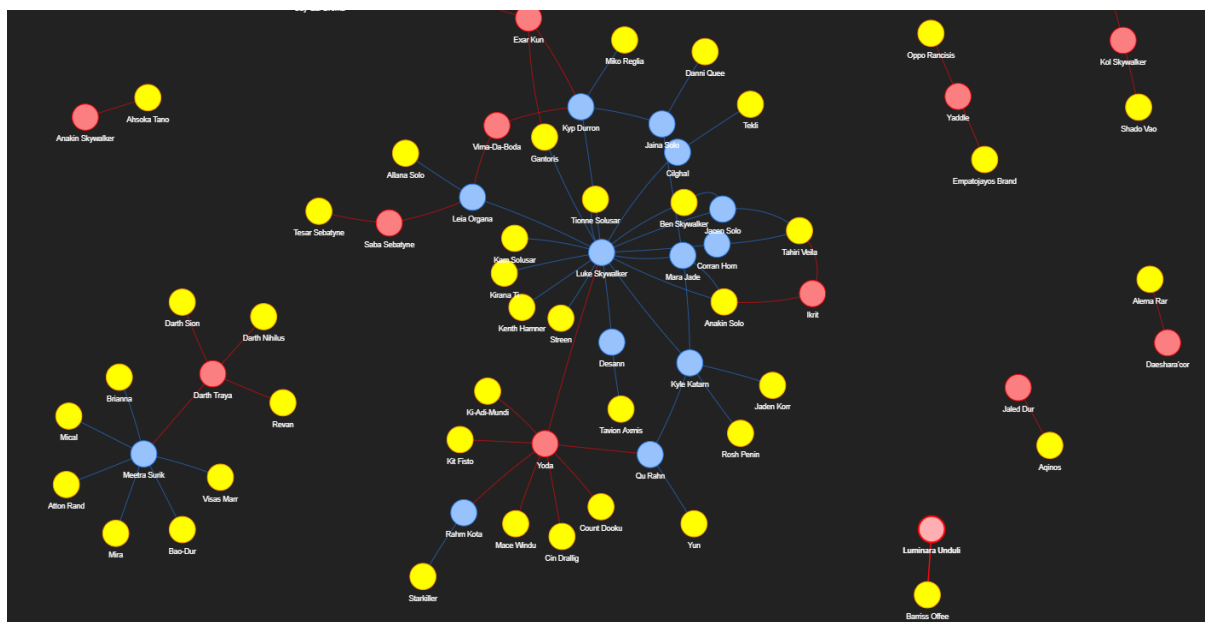


Quels sont les relations maître / padawan entre les Jedi ?

Résultat de la requête au format csv :

Student	Master
Kyp Durrón	Luke Skywalker
Tionne Solusar	Luke Skywalker
Allana Solo	Leia Organa
Ulic Qel-Droma	Exar Kun
Darth Sion	Darth Traya
Ki-Adi-Mundi	Yoda
Cin Drallig	Yoda
Danni Quee	Jaina Solo

Représentation sous forme de graphe :



Analyse des résultats :

Le code couleur du graphe est le suivant : les individus en rouge sont des maîtres, les individus en jaune des élèves et les individus bleus ont eu les 2 rôles au cours de leur vie.

On peut ainsi voir qu'un certain nombre de personnages gravitent autour des membres principaux de la franchise comme Luke Skywalker ou Yoda. Cela est entre autres expliqué par le nombre plus important d'œuvres incluant ces personnages.

On retrouve aussi un certain nombre de binômes maîtres/élève, souvent introduit et seulement présent dans l'univers étendu de la série (jeux et bande dessinés).