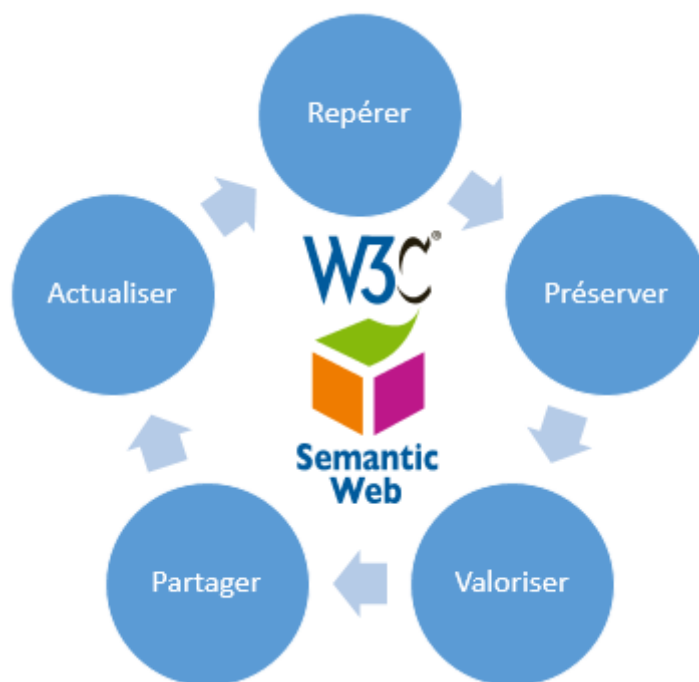


Travail de Bachelor 2018

Search engine optimization (SEO) au moyen de technologies sémantiques



Déposé le 8 août 2018

Etudiant : Jonathan Joaquim

Professeur : Maria Sokhn

Source de l'illustration sur la page de titre

<https://www.adbs.fr/sites/default/files/styles/body/public/ill-web-sem-cycle-gestion-des-connaissances-et-web-semantique.png?itok=suVtNTzh>

I. Résumé

L'entreprise Klewel est spécialisée dans la rediffusion d'événement sur le web. Leur solution permet l'enregistrement d'événement, de conférence ou même de formation. Ceux-ci sont ensuite sauvegardés et pourront être exploités par le client. Le tout pouvant être intégré au site internet du client, en interne ou sur le portail web de Klewel.

L'objectif du présent travail de Bachelor est d'optimiser la recherche de ces conférences sur le portail de Klewel via les moteurs de recherche. Pour cela, deux aspects du Web sémantique ont été réalisés :

- Améliorer la visibilité du site grâce aux micro-données avec les notions de schema.org
- Implémenter un sitemap.xml pouvant être fourni aux moteurs de recherche

Mots-clés : web sémantique, micro-données, sitemap, RDF, webcasting

II. Avant-propos

Le document présente le travail effectué dans le cadre du travail de Bachelor en filière Informatique de Gestion à la Haute-Ecole spécialisée Suisse occidentale (HES-SO).

Ce sujet proposé m'a intéressé car le web sémantique est une nouvelle grande étape du Web. Je trouvais donc important d'apprendre son utilité et comment exploiter les informations transmises aux différents moteurs de recherche.

III. Remerciements

Je remercie les personnes suivantes pour leur aide et leur soutien dans la réalisation de ce travail de Bachelor :

Maria Sokhn, pour son soutien, sa grande disponibilité et son optimisme

Samuel Dorsaz, pour son aide et son temps investi

Fabian Cretton, pour ses conseils

IV. Table des matières

Glossaire	9
Introduction.....	10
1 Gestion du projet	12
1.1 Planification	12
1.2 Initialisation.....	12
1.3 Analyse.....	12
1.4 Implémentation	12
1.5 Finalités	12
2 Prise en main du contexte actuelle de Klewel	13
3 Etat de l'art	14
3.1 Le Web sémantique.....	14
3.2 Schéma.org	15
3.3 RDF.....	16
3.4 RDFa.....	17
3.5 JSON-LD	17
3.6 Le sitemap.....	18
4 Choix et prise de position.....	19
4.1 Analyse de la complexité de mise en œuvre.....	19
4.1.1 Micro-données.....	19
4.1.2 Sitemap.xml.....	19
4.2 Proposition de la structure la plus adaptée	19
4.2.1 Micro-données.....	19
4.2.2 Sitemap.xml.....	20
5 Implémentation	21
5.1 Architecture de l'application	21
5.2 Micro-données avec schéma.org	22
5.2.1 Implémentation sur la page html webcast.....	23
5.2.2 Implémentation sur la page html talk	27

5.2.3	Test	30
5.3	Sitemap.xml avec le framework Django	34
5.3.1	Sitemap des conférences	34
5.3.2	Sitemap des talks	37
	Conclusion.....	39
	Bilan du travail.....	39
	Idées d'amélioration et perspectives	39
	Références	40
	Annexes.....	42
I.	Diagramme de Gantt.....	42
II.	Contenu du CD.....	43
	Déclaration de l'auteur	44

VI. Liste des tableaux

Tableau 1 : Différence entre RDFa et JSON-LD (données de l'auteur)	19
--	----

VII. Liste des figures

Figure 1 : Illustration du player sur le portail klewel (Guillemot)	10
Figure 2 : Représentation de la planification (données de l'auteur)	12
Figure 3 : Fichier host du répertoire Windows/System32 (données de l'auteur)	13
Figure 4 : Propriétés disponibles pour l'élément Event (Event, s.d.)	16
Figure 5 : Balise html sans marquage (Person, s.d.)	17
Figure 6 : Même balise qu'en figure 5 avec marquage RDFa (Person, s.d.)	17
Figure 7 : Script JSON-LD des informations présentes en figure 5 (Person, s.d.)	18
Figure 8 : Simplification du diagramme de classe (données de l'auteur)	21
Figure 9 : Session privée connectée à tous les talks (données de l'auteur)	22
Figure 10 : Session publique avec uniquement deux talks (données de l'auteur)	22
Figure 11 : Exemple de conférence sur le portail de Klewel en serveur test (données de l'auteur)	23
Figure 12 : Section de la page HTML webcast où le RDFa a été implémenté (données de l'auteur)	24
Figure 13 : Résultat du RDFa dans le code source de la page présentée en figure 11 (données de l'auteur)	25
Figure 14 : Script JSON-LD implémenté dans la page HTML webcast (données de l'auteur)	26
Figure 15 : Résultat du script dans le code source de la page présentée en figure 11 (données de l'auteur)	26
Figure 16 : Exemple de talk sur le portail de Klewel en serveur test (données de l'auteur)	27
Figure 17 : Section de la page HTML talk où le RDFa a été implémenté (données de l'auteur)	28
Figure 18 : Résultat du RDFa dans le code source de la page présentée en figure 16 (données de l'auteur)	28
Figure 19 : Script JSON-LD implémenté dans la page HTML talk (données de l'auteur)	29

Figure 20 : : Résultat du script dans le code source de la page présentée en figure 16 (données de l'auteur)	30
Figure 21 : Résultat du RDFa en figure 13 sur l'outil de Google. (Outil de test des données structurées, s.d.).....	31
Figure 22 : Résultat du JSON-LD en figure 15 sur l'outil de Google. (Outil de test des données structurées, s.d.)	32
Figure 23 : Résultat du RDFa en figure 18 sur l'outil test de Google.....	33
Figure 24 : Résultat du JSON-LD en figure 20 sur l'outil test de Google.....	33
Figure 25 : Extrait du code implémenter dans la classe Conférence (données de l'auteur) ..	34
Figure 26 : Fonction permettant de définir l'URL de la conférence (données de l'auteur)	35
Figure 27 : Liste des redirections selon la zone du site voulu (données de l'auteur).....	36
Figure 28 : Résultat du sitemap.xml des conférences (Klewel, s.d.)	37
Figure 29 : Exemple graphique d'une session privée (données de l'auteur)	38
Figure 30 : Exemple d'URL d'un talk (données de l'auteur)	38
Figure 31 : Exemple graphique d'une session publique (données de l'auteur).....	38

Glossaire

Termes	Définition
Django	Framework de python
HTML	HyperText Markup Language – langage utilisé pour structurer une page web (HTML (HyperText Markup Language) : définition, traduction, 2018)
Micro-données	Permet d'ajouter du contenu sémantique dans des documents HTML. (Briche, 2018)
Schéma.org	Schéma de micro-données utilisé par le Web. (Schema.org : définition, 2018)
RDF	Ressource Description Framework – modèle utilisé pour le référencement des ressources Web et leurs métadonnées. (Rouse, 2005)
RDFa	RDF Attributs – syntaxe pour structurer les données dans une page HTML (Bathelot, Définition : RDFa, 2015)
JSON-LD	JavaScript Object Notation Linked Data – syntaxe pour encoder des données structurées grâce à du JSON (Vallette, 2015).
URI	Uniform Resource Identifier – format d'identité pour une ressource sur un réseau internet (Quelle est la différence entre URI, URL et URN ?, 2018)
URL	Uniform Resource Locator – format de nommage universel désignant une ressource d'internet (Quelle est la différence entre URI, URL et URN ?, 2018)

Introduction

L'entreprise Klewel SA, basée à Martigny, possède un portail web permettant à ces clients de déposer l'enregistrement de leurs conférences. Celles-ci sont ensuite disponibles de manière publique ou privée afin d'être revisualisée. La particularité de ce portail, qui est aussi la force de cette entreprise, est la navigation au travers de la conférence.

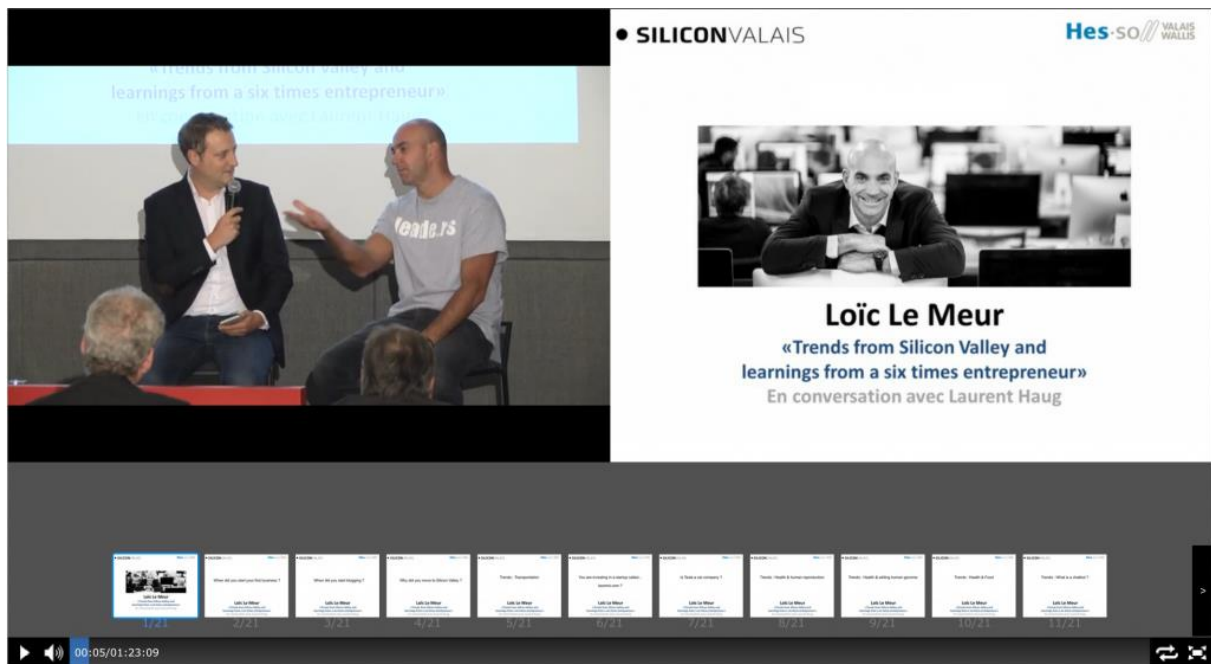


Figure 1 : Illustration du player sur le portail klewel (Guillemot)

La visualisation des différents slides permet de naviguer dans la présentation. Par exemple, si seul le contenu de la cinquième diapositive intéresse l'utilisateur, il peut cliquer dessus et la vidéo s'ajustera directement sur le moment concernant cette diapositive. Klewel fournit aussi pour certaines vidéos, des sous-titres pouvant être affichés sur le player de leur portail (Klewel, s.d.).

L'entreprise Klewel possède beaucoup de données et elle aimerait en améliorer leur référencement auprès des moteurs de recherches. Le Web sémantique est la solution afin d'exploiter toutes ces données. Grâce à plusieurs méthodes, les informations peuvent être triées et référencées afin que chaque donnée ne soit pas simplement du texte mais plutôt une définition.

Par exemple, si nous recherchons la couleur orange mais n'écrivons que « orange », un moteur de recherche ressortira toutes les pages parlant soit de la couleur, soit du fruit et ce simplement en trouvant le mot orange sur la page web. Même en écrivant « couleur orange », la liste des résultats de recherche possèdera quand même des pages parlant de

fruits alors que cela ne nous intéresse pas. Grâce au Web sémantique, le mot orange peut être référencié comme une couleur, ou comme le fruit. Les moteurs de recherche sont alors capables d'exploiter ces informations et d'éviter tout malentendu dans la liste des résultats et permettent d'offrir une réponse optimale à l'utilisateur.

En plus d'aider l'utilisateur à trouver plus facilement ce qu'il recherche, le Web sémantique permet de supprimer de la liste des résultats de recherche des pages ne concernant pas vraiment ce qui est recherché. Permettant aux sites qui peuvent réellement intéresser l'utilisateur d'apparaître plus facilement dans la liste et donc d'augmenter leur taux de visites. De nos jours, c'est plus de quinze millions de sites utilisant le Web sémantique.

« Les gens demandent tout le temps ce qu'est le Web 3.0. Je suppose que lorsque vous aurez une superposition de dessins vectoriels — toute en vagues et en plis brumeux — décrivant le Web 2.0 et l'accès à un Web sémantique intégré dans un immense espace de données, vous aurez accès à une incroyable ressource de données. » (Shanon, 2006)

Cette phrase prononcée par Tim Berners-Lee, l'inventeur du web dans les années 90 et fondateur du World Wide Web Consortium en 1996, démontre l'importance du web sémantique.

1 Gestion du projet

Ce projet n'a pas suivi de technique de gestion stricte. Un planning avec plusieurs étapes a toutefois été réalisé afin d'avoir une ligne de conduite tout au long du travail.



Figure 2 : Représentation de la planification (données de l'auteur)

1.1 Planification

Pour compléter ce schéma, un diagramme de Gantt complet est fourni en annexe.

1.2 Initialisation

La première étape fut de prendre en main la machine virtuelle fournie directement par Klewel. Plutôt que de fournir simplement des concepts à Klewel, pouvoir travailler directement sur leur serveur test a permis d'implémenter directement le Web sémantique dans leur application. L'entreprise pourra donc rapidement exploiter ce travail.

1.3 Analyse

En second, il était important de clarifier ce qu'était le web sémantique et pourquoi est-ce si important de nos jours. Un autre obstacle fut l'appréhension du langage Python et de son framework Django. Pour terminer, il a fallu choisir une méthode d'implémentation.

1.4 Implémentation

La partie principale de ce travail était d'implémenter le web sémantique à l'application de Klewel ainsi que l'écriture de ce travail.

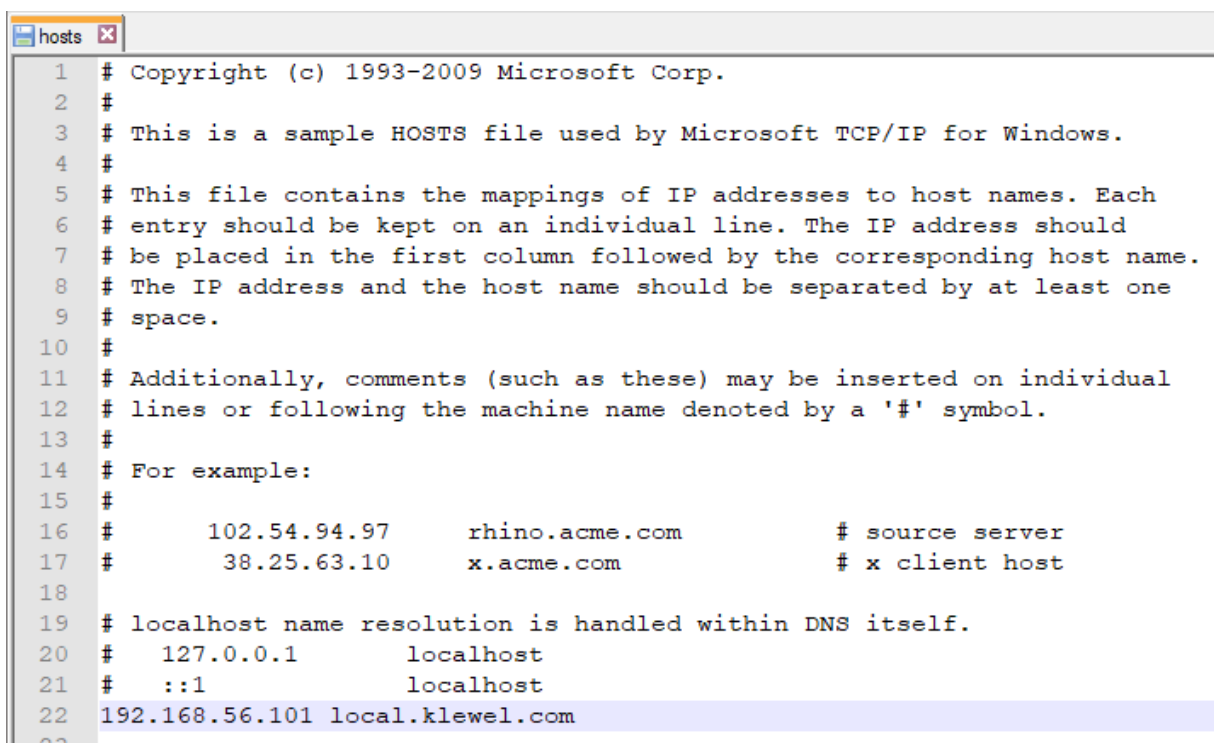
1.5 Finalités

La conclusion de ce travail était le contrôle de l'orthographe ainsi que l'anti-plagiât.

2 Prise en main du contexte actuelle de Klewel

L'entreprise Klewel a fourni une machine virtuelle permettant de se connecter en local à un serveur test de leur portail vidéo. Grâce à cela, il a été possible d'implémenter le code nécessaire directement dans leur application.

L'utilisation de cette VM (Virtual Machine) nécessitait l'application Oracle VM Virtual Box. Pour avoir accès à l'adresse « local.klewel.com », il a été nécessaire d'ajouter au fichier système Windows « host » une ligne indiquant à quelle adresse IP elle faisait référence.

A screenshot of a Windows hosts file. The window title is 'hosts'. The file contains several lines of text, mostly comments starting with '#'. Line 22 is highlighted in blue and contains the entry '192.168.56.101 local.klewel.com'.

```
1 # Copyright (c) 1993-2009 Microsoft Corp.
2 #
3 # This is a sample HOSTS file used by Microsoft TCP/IP for Windows.
4 #
5 # This file contains the mappings of IP addresses to host names. Each
6 # entry should be kept on an individual line. The IP address should
7 # be placed in the first column followed by the corresponding host name.
8 # The IP address and the host name should be separated by at least one
9 # space.
10 #
11 # Additionally, comments (such as these) may be inserted on individual
12 # lines or following the machine name denoted by a '#' symbol.
13 #
14 # For example:
15 #
16 #       102.54.94.97       rhino.acme.com       # source server
17 #       38.25.63.10       x.acme.com           # x client host
18
19 # localhost name resolution is handled within DNS itself.
20 #   127.0.0.1       localhost
21 #   ::1             localhost
22 192.168.56.101 local.klewel.com
23
```

Figure 3 : Fichier host du répertoire Windows/System32 (données de l'auteur)

Une fois la machine virtuelle lancée, un outil de connexion ssh devait être utilisé pour se connecter à `ubuntu@local.klewel.com`. Il ne restait plus qu'à préciser où se trouvait le code de l'application pour que les deux soient reliés.

3 Etat de l'art

Avant d'implémenter le Web sémantique, nous avons plusieurs notions importantes à définir et comprendre. Nous devons aussi analyser les différentes technologies et méthodes disponibles afin d'intégrer le web sémantique à l'application.

3.1 Le Web sémantique

Le Web sémantique est un standard web introduit par le W3C. Le but est d'encourager l'échange des données grâce à des technologies établies. Comme l'a dit son inventeur Tim Berners-Lee, l'idée est de passer d'un Web de données simplement stockées à un Web dit « intelligent » (Web sémantique, web de données : définitions, 2018). Grâce à ces formats, les données pourraient être réutilisées par les ordinateurs afin d'apporter à l'utilisateur ce qu'il recherche vraiment plutôt qu'une liste de possibilités. (Comprendre le web sémantique, s.d.)

Mais malgré sa grande utilité, quelques préoccupations limitent son expansion à la totalité du Web actuel. Contrairement à l'humain, l'ordinateur ne peut mettre en doute une information qui lui est transmise. De ce fait, toute donnée faussée étant fournie pourra totalement compromettre le résultat. Dans des domaines tels que la science ou encore la médecine, les répercussions d'avoir des données altérées, voir erronées, pourraient être désastreuses.

Un autre frein à son développement est la confidentialité. Pour que les données puissent être utilisées entre application, il est logique que chacun fournisse lui aussi des données. Malheureusement la réalité est autre. Les plus grandes entreprises informatiques comme Google, Apple ou encore Facebook n'ont jamais été enclins à révéler leurs données. Mais à l'inverse, elles sont connues pour exploiter le plus grand nombre de données possibles (Maignien, 2014).

Afin d'exploiter ce Web sémantique, le W3C a créé plusieurs protocoles afin que des standards puissent être utilisés par les sites. Deux protocoles ont été utilisés dans ce travail : Le RDFa et JSON-LD. Afin de fournir un maximum de données et de possibilités à Klewel, les deux méthodes ont été implémentées. Il est toutefois important de n'en utiliser qu'une dans l'application en production ou les données seront doublées et le moteur de recherche analysera deux éléments différents alors qu'il n'y en a qu'un seul.

3.2 Schéma.org

Schéma.org est une initiative de groupe lancée par Google, Bing et Yahoo en 2011. Il représente un vocabulaire pouvant être utilisé pour catégoriser les données (Schema.org : définition, 2018).

Malheureusement, ce vocabulaire est encore incomplet. Par exemple, il n'est pas encore possible de détailler les animaux ou les insectes. Les termes disponibles sont toutefois suffisants pour l'élaboration de ce travail.

Chaque terme disponible sur schéma.org possède des propriétés. Par exemple, le terme « Person » possède des propriétés tels que nom, genre, nationalité, date de naissance et bien d'autres. Ce sont ces différents mots que nous allons lier aux éléments présents sur les pages HTML afin que l'ordinateur puisse comprendre quel élément désigne quelle propriété et ensuite transmettre ces données aux moteurs de recherche.

Schema.org permet de connecter différents mots de vocabulaire. Pour rester dans les notions qui seront utilisées durant ce travail, prenons l'exemple d'un évènement, plus précisément une conférence. La conférence pourra être identifiée comme élément « Event » disponible sur schema.org. Mais cet événement a bien lieu quelque part sur terre, cette localisation utilisera l'attribut « PostalAddress ». Une conférence possède aussi toujours au moins une personne en tant que présentateur. Elle pourra être défini avec l'élément « Person ».

Property	Expected Type	Description
Properties from Event		
about	Thing	The subject matter of the content. Inverse property: subjectOf .
actor	Person	An actor, e.g. in tv, radio, movie, video games etc., or in an event. Actors can be associated with individual items or with a series, episode, clip. Supersedes actors .
aggregateRating	AggregateRating	The overall rating, based on a collection of reviews or ratings, of the item.
attendee	Organization or Person	A person or organization attending the event. Supersedes attendees .
audience	Audience	An intended audience, i.e. a group for whom something was created. Supersedes serviceAudience .
composer	Organization or Person	The person or organization who wrote a composition, or who is the composer of a work performed at some event.
contributor	Organization or Person	A secondary contributor to the CreativeWork or Event.
director	Person	A director of e.g. tv, radio, movie, video gaming etc. content, or of an event. Directors can be associated with individual items or with a series, episode, clip. Supersedes directors .
doorTime	DateTime	The time admission will commence.
duration	Duration	The duration of the item (movie, audio recording, event, etc.) in ISO 8601 date format.
endDate	Date or DateTime	The end date and time of the item (in ISO 8601 date format).
eventStatus	EventStatusType	An eventStatus of an event represents its status; particularly useful when an event is cancelled or rescheduled.
funder	Organization or Person	A person or organization that supports (sponsors) something through some kind of financial contribution.
inLanguage	Language or Text	The language of the content or performance or used in an action. Please use one of the language codes from the IETF BCP 47 standard. See also availableLanguage . Supersedes language .
isAccessibleForFree	Boolean	A flag to signal that the item, event, or place is accessible for free. Supersedes free .
location	Place or PostalAddress or Text	The location of for example where the event is happening, an organization is located, or where an action takes place.

Figure 4 : Propriétés disponibles pour l'élément Event (Event, s.d.)

La figure ci-dessus représente une partie des informations d'un évènement pouvant être envoyé aux moteurs de recherche. Et la partie « Expected type » démontre que la plupart des éléments sont directement connecter comme un autre terme du vocabulaire de schema.org. C'est ce qui rend cette sémantique aussi performante.

3.3 RDF

Afin que les données puissent être utilisées, il était primordial de créer un modèle de représentation. RDF est la solution inventée par le W3C. Sous la forme d'un triplet, l'information est stockée en trois parties : Le sujet, le prédicat et l'objet.

Par exemple, « L'inventeur du Web sémantique est Tim Berners-Lee », le Web sémantique correspond au sujet, le prédicat est le mot inventeur et l'objet Tim Berners-Lee.

Un élément essentiel pour cette transmission des données est que chaque donnée possède un URI (Uniform Resource Identifier) ressemblant à des URL. Chaque URI étant unique, on peut donc connecter chaque donnée les unes aux autres sans risque de mauvaises informations (Rouse, 2005).

3.4 RDFa

Dans une page HTML, l'être humain possède une qualité d'analyse que les ordinateurs ne possèdent pas à l'heure actuelle. L'utilisateur saura faire la différence entre un titre, entre un menu d'aide à la navigation ou encore un simple paragraphe de texte. Pour les machines, ce n'est que du texte (Adida & Birbeck, 2008) (Cuvelier, 2011).

Le RDF in HTML attributes (RDFa) permet, comme son nom l'indique, d'insérer aux balises HTML des attributs permettant de préciser quel est le type de donnée.

`<p>Christopher Froome was sponsored by Sky in the Tour de France.</p>`

Figure 5 : Balise html sans marquage (Person, s.d.)

```
<p vocab="http://schema.org/" typeof="Person">
  <span property="name">Christopher Froome</span> was sponsored by
  <span property="sponsor" typeof="http://schema.org/Organization">
    <a property="url" href="http://www.skysports.com/">Sky</a></span> in the Tour de France.
</p>
```

Figure 6 : Même balise qu'en figure 5 avec marquage RDFa (Person, s.d.)

Grâce à ce marquage, l'ordinateur peut analyser que le texte parle d'une personne. Son nom étant Christopher Froome et son sponsor, une organisation s'appelant SkySports.

3.5 JSON-LD

JSON est utilisé pour le formatage des données en texte lisible. Il permet de représenter des objets ou des structures de données dans des navigateurs Web (Vallette, 2015).

JSON-Linked Data (JSON-LD) est une autre méthode que le RDFa appuyée par le W3C. Contrairement à RDFa, JSON-LD est intégré grâce à script caché dans la page Web. Il n'est donc pas nécessaire de marquer les balises HTML.

```
<script type="application/ld+json">
{
  "@context": "http://schema.org/",
  "@type": "Person",
  "name": "Christopher Froome",
  "sponsor":
  {
    "@type": "Organization",
    "name": "Sky",
    "url": "http://www.skysports.com/"
  }
}
</script>
```

Figure 7 : Script JSON-LD des informations présentes en figure 5 (Person, s.d.)

3.6 Le sitemap

Un plan de site est un fichier XML fournissant aux moteurs de recherche une indexation d'un site Web. Il permet de hiérarchiser l'ensemble des pages du site concerné. L'avantage d'avoir un sitemap est d'intégrer des informations supplémentaires pour chaque page. Par exemple, la date de dernière modification ou encore son importance par rapport à d'autres pages du site. De ce fait, le moteur de recherche peut naviguer sur le site de manière plus intelligente (Bathelot, Définition : Sitemap, 2017).

4 Choix et prise de position

4.1 Analyse de la complexité de mise en œuvre

4.1.1 Micro-données

L'avantage de pouvoir travailler directement sur l'application de Klewel fut la possibilité d'analyser directement les pages HTML du site. Leur portail vidéo étant déjà bien abouti, la recherche des différents termes de vocabulaire pouvant être intégré fut grandement facilité.

4.1.2 Sitemap.xml

Le framework Django offre plusieurs possibilités afin de créer automatiquement un sitemap. La complexité d'implémentation repose ici sur l'architecture de la base de données. Afin de générer ce plan de site, il est nécessaire de préciser pour chaque page son url. Comme cela va être expliqué dans le chapitre 7.3, la structure de l'application limite l'utilisation des différentes solutions fournies par Django.

4.2 Proposition de la structure la plus adaptée

4.2.1 Micro-données

Pour les besoins de cette application, deux choix correspondent : Le RDFa ou le JSON-LD.

	RDFa	JSON-LD
Implémentation	Dans les balises HTML concernées	Script indépendant
Visibilité	Nécessaire de chercher dans chaque balise	Facilement lisible
Double données	Non	Oui

Tableau 1 : Différence entre RDFa et JSON-LD (données de l'auteur)

Bien que les deux possibilités aient été implémentées pour ce travail. Le RDFa conviendra mieux pour éviter la répétition des données. Dans l'exemple de Klewel, comme expliqué dans l'introduction, des sous-titres peuvent être présents pour une vidéo. Ceux-ci seront intégrés à la page web et peuvent donc générer une quantité de mots importante à charger. Les sous-titres pouvant être utilisés comme information pour le Web sémantique, la

solution du JSON-LD n'est pas adaptée car elle requerrait de les charger à double sur la page.

4.2.2 Sitemap.xml

Le langage Python ainsi que le framework Django proposent différents packages afin d'intégrer à l'application une auto-documentation du sitemap.xml. Afin de garantir son fonctionnement, il est plus sûr d'utiliser un package de Django que simplement de Python.

Le framework Django possède une fonction pour les sitemaps déjà intégré. Il est donc conseillé de l'utiliser. Le développeur peut toutefois trouver d'autres packages, tel que `django-qarte` ou `django-feedmap`. Ces packages offrent d'autres méthodes afin d'intégrer un sitemap.

La fonction déjà intégrée étant la plus développée, documentée et surtout tenue à jour, elle sera utilisée dans cette application.

Il est important de noter que lors de la réception de l'application pour accomplir ce travail de Bachelor, un « Proof of concept » avait déjà été intégré par Samuel Dorsaz, chef de projet chez Klewel. Cette base implémentée afin de vérifier la faisabilité d'un sitemap.xml fût gardée car elle utilisait la fonction intégrée de sitemap et était parfaitement fonctionnelle. Il sera démontré dans le chapitre 7.3 ce qui était déjà présent dans l'application et ce qui a été implémenté.

5 Implémentation

Afin de comprendre les termes utilisés par l'entreprise, il est important de visualiser les connexions entre les éléments de leurs bases de données. Une fois ce vocabulaire assimilé, l'implémentation peut être effectuée.

5.1 Architecture de l'application

L'application Klewel possède une base de données avec plus de 20 tables. La figure ci-dessous représente une version simplifiée des tables essentielles à ce projet.

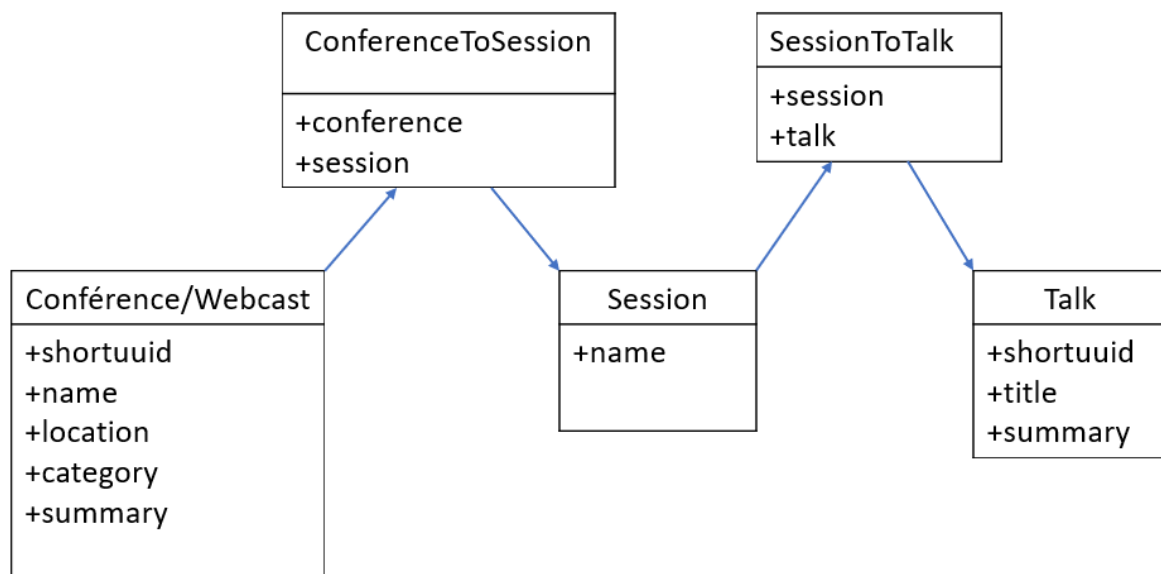


Figure 8 : Simplification du diagramme de classe (données de l'auteur)

Une conférence, qui correspond sur le portail klewel à un webcast, représente l'événement. Elle possède des informations essentielles telles que la localisation, une catégorie, par exemple l'environnement et un résumé.

Le talk représente une intervention durant cette conférence. Si l'on reste sur le thème de l'environnement, un talk parlera du réchauffement climatique, un autre de la fonte des glaces. Il est important de comprendre que ces interventions peuvent être rendues publiques ou privées sur le portail de Klewel. Par exemple, si un client est une entreprise, la conférence peut-être un événement interne. Certaines ou même toutes les interventions peuvent donc être confidentielles et seuls les employés de l'entreprise devront avoir accès aux vidéos sur le portail.

L'élément de session est une particularité de Klewel pour répondre à ce besoin de confidentialité. Si durant cette conférence interne à l'entreprise six talks avait eu lieu, le client pourrait choisir d'en rendre deux publiques et les quatre autres seraient privées. Dans ce

cas, la base de données Klewel créerait deux sessions, la première avec uniquement les deux interventions publiques et la seconde avec la totalité des talks en privés. La notion de Session n'étant pas directement visible sur le portail étant donné que seul les talks publics

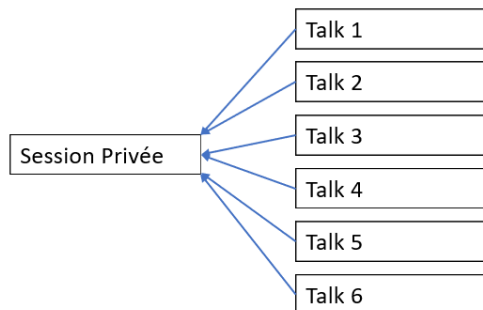


Figure 9 : Session privée connectée à tous les talks (données de l'auteur)

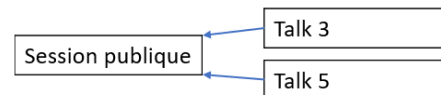


Figure 10 : Session publique avec uniquement deux talks (données de l'auteur)

y sont affichés.

Cette particularité oblige la base de données à avoir une relation dit de « plusieurs à plusieurs » entre la table Session et la table Talk.

5.2 Micro-données avec schéma.org

La première étape pour implémenter les micro-données était de trouver quels termes disponibles sur schéma.org convenaient le mieux à une conférence. Le terme « Event » englobe tout type d'événement comme par exemple des festivals, des manifestations sportives ou encore une conférence sur l'environnement. L'idée d'utiliser la sous-catégorie « BusinessEvent » a été abordée avec Klewel mais après discussion il s'avère que le portail n'accueille pas uniquement des conférences professionnelles mais aussi sur d'autres sujets tels que la santé ou le sport.

Mais si la conférence est un « Event », il faut donc trouver une terminologie plus précise pour le talk. L'idée étant de ne pas utiliser les sous-catégories de « Event » car le développeur aurait besoin de modifier la sous-catégorie pour chaque événement. L'objectif de ce travail est d'éviter aux développeurs de Klewel de devoir ajuster le code et que toutes les informations soient englobées. Pour cette raison, le terme « Event » correspondait mieux à un talk qu'à une conférence.

Il nous fallait donc trouver un terme plus large afin de correspondre à une conférence. Le terme « EventSeries » correspondait parfaitement. Comme son nom l'indique, il désigne une série d'événements ayant lieu à un même endroit. La page Web concernant la conférence

informera ainsi le moteur de recherche que plusieurs interventions sont liées à celle-ci et chaque page concernant un talk précisera faire partie de cet « EventSeries ».

5.2.1 Implémentation sur la page html webcast





Comme présenté dans la figure ci-dessous, il y a plusieurs informations pouvant être envoyées aux moteurs de recherche. Par exemple, le nom de la conférence, sa localisation, la date, un résumé. Même le logo qui est un lien vers le site de l'entreprise responsable de cette conférence peut être lié. Dans ce cas précis, une simple recherche de l'entreprise Ferring permettra au lien vers le portail de Klewel d'apparaître dans les résultats.

Webcasts / Tailoring mild-to-moderate IBD management

Tailoring mild-to-moderate IBD management

Amsterdam, The Netherlands · mars 2016 · 4 Talks

Patient-reported outcomes are becoming an increasingly important part of research and clinical management, but how can we integrate them with established clinical measures of disease activity? It gives us great pleasure to invite you to this Ferring-sponsored symposium. We shall explore mild-to-moderate Crohn's disease and ulcerative colitis case studies in the context of guidelines and clinical evidence as well as how patient-reported outcomes can help to align patient priorities with clinical practice, allowing you to tailor treatments to the needs of your patients.

Share:    

Conference program





 <p>How to combine PROs and guidelines in your clinical practice: Illustrated with case studies. Ulcerative colitis Axel Dignass, Agaplesion Markus Hospital, Frankfurt, Germany 18 mars 2016 · 12:51 après-midi 273 views</p>	 <p>Panel Discussion 19 mars 2016 · 1:02 après-midi 298 views</p>
 <p>Patient-reported outcomes (PROs) - What does this mean? Silvio Danese, Humanitas Research Hospital, Milan, Italy 20 mars 2016 · 12:31 après-midi 191 views</p>	 <p>How to combine PROs and guidelines in your clinical practice: Illustrated with case studies. Crohn's disease Edouard Louis, University Hospital CHU Liège, Belgium 21 mars 2016 · 12:45 après-midi</p>

Figure 11 : Exemple de conférence sur le portail de Klewel en serveur test (données de l'auteur)

5.2.1.1 Implémentation du RDFa

Comme indiqué, la lisibilité n'est pas optimale, sans connaissances basiques en HTML et Web sémantique, il peut être difficile de trouver les parties concernant le RDFa.

```
<div class="span12" vocab="http://schema.org/" typeof="EventSeries">  
  <div class="page-header" style="border-style: none; margin-bottom: 10px;">  
    <h1><span property="name">  
      {{ webcast.name }}</span><br/>  
      <small>  
        {% if webcast.location %}  
          <span property="location" typeOf="Place"><span property="address">  
            {{webcast.location }}</span></span>&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&~  
          {% endif %}  
          <span property="startDate" element={{ webcast.start_date | date:"Y-m-d" }}>  
            {{ webcast.start_date | date:"F Y" }}</span>&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&~  
            {{ webcast.getNumTalks }}&nbsp;&nbsp;& Talk{{ webcast.getNumTalks | pluralize }}  
          </span property="about" element="{{webcast.category}}">  
        </small>  
        {% if webcast.logo.getURL %}  
          <span property="image" typeOf="ImageObject" >  
            <a href="{{ webcast.logo.link }}" property="url" typeOf="URL" target="_blank">  
                
            </a>  
          </span>  
          {% endif %}  
        </h1>  
      </div>  
      {% if webcast.summary %}  
        <blockquote style="margin-bottom: 40px; text-align: justify;">  
          <p property="description">{{ webcast.summary }}</p>  
        </blockquote>  
      {% endif %}  
    </div>
```

Figure 12 : Section de la page HTML webcast où le RDFa a été implémenté (données de l'auteur)

Un avantage que partagent les différents standards pouvant être utilisés est l'intégration d'information à envoyer aux moteurs de recherche qui ne seront pas forcément affichés. Par exemple, dans la figure 11, nous pouvons constater que la catégorie de la conférence, qui peut être aussi définie comme thème, n'est pas précisée sur la page. C'est pourtant une information importante envoyée aux moteurs de recherche. Il est donc possible d'intégrer d'autres informations sans forcément les faire apparaître. C'est le cas dans la figure 12 : La ligne « `` » n'affiche aucun élément sur la page Web. Pourtant elle permettra d'envoyer la catégorie de la conférence aux moteurs de recherche.

[illegible]

Figure 13 : Résultat du RDFa dans le code source de la page présentée en figure 11 (données de l'auteur)

Le résultat nous permet de constater que la catégorie de la conférence est bien référencée même si elle n'apparaît pas sur la page.

5.2.1.2 Implémentation JSON-LD

L'avantage que possède JSON-LD sur le RDFa est qu'il s'agit d'un script isolé. Il est donc plus simple à implémenter et à lire.

```

<script type="application/ld+json">
{
  "@context" : "http://schema.org",
  "@type" : "EventSeries",
  "name": "{ webcast.name }",
  "about": "{ webcast.category }",
  "startDate": "{ webcast.start_date|date:'Y-m-d' }",
  "description": "{ webcast.summary }",

  "location":
  {
    "@type": "Place",
    "address" :
    {
      "@type": "PostalAddress",
      "name": "{ webcast.location }"
    }
  },
  "image":
  {
    "@type": "ImageObject",
    "url": "{ webcast.logo.link }"
  }
}
</script>

```

Figure 14 : Script JSON-LD implémenté dans la page HTML webcast (données de l'auteur)

Les informations fournies dans ce script sont exactement les mêmes que dans les balises pour le RDFa. Bien que le script soit visible dans le code source, aucune information ne sera affichée sur la page Web.

```

<script type="application/ld+json">
{
  "@context" : "http://schema.org",
  "@type" : "EventSeries",
  "name": "Tailoring mild-to-moderate IBD management",
  "about": "Health Sciences",
  "startDate": "2016-03-18",
  "description": " Patient-reported outcomes are becoming an increasingly important part of research and clinical
management, but how can we integrate them with established clinical measures of disease activity? It gives us great pleasure to
invite you to this Ferring-sponsored symposium. We shall explore mild-to-moderate Crohn's disease and ulcerative colitis case
studies in the context of guidelines and clinical evidence as well as how patient-reported outcomes can help to align patient
priorities with clinical practice, allowing you to tailor treatments to the needs of your patients.      ",

  "location":
  {
    "@type": "Place",
    "address" :
    {
      "@type": "PostalAddress",
      "name": "Amsterdam, The Netherlands"
    }
  },
  "image":
  {
    "@type": "ImageObject",
    "url": "http://www.ferring.com/"
  }
}
</script>

```

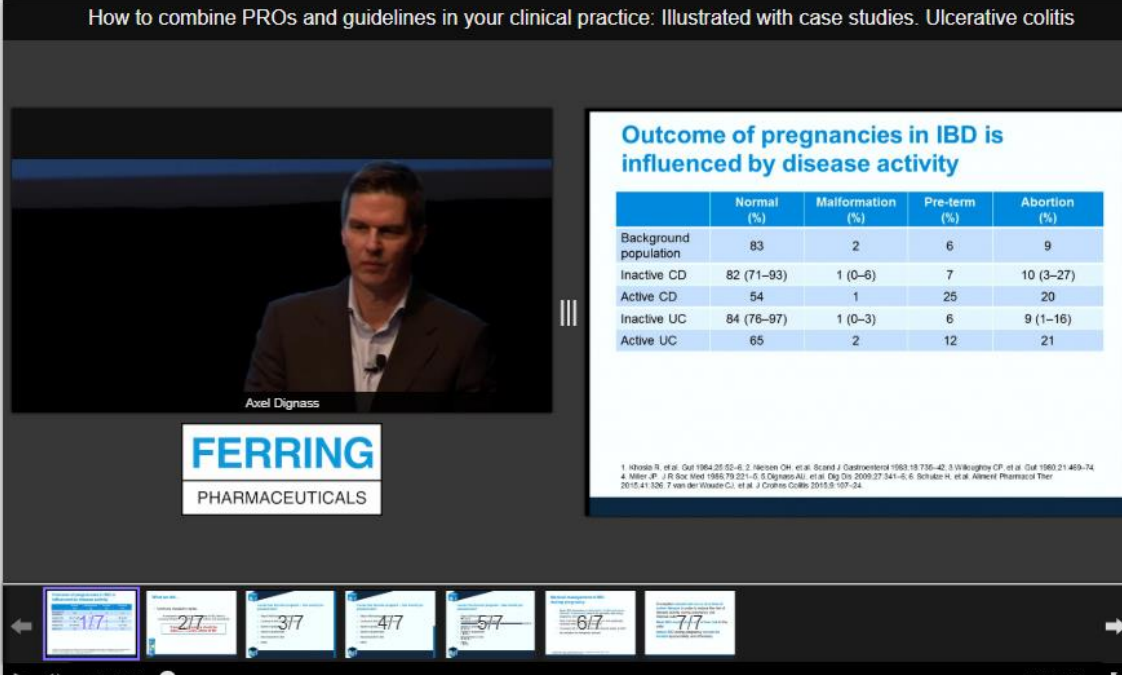
Figure 15 : Résultat du script dans le code source de la page présentée en figure 11 (données de l'auteur)

Nous pouvons constater dans la figure 15 la lisibilité de ce script. Mais contrairement au RDFa, ces informations sont un double de ce qui est affiché sur la page. Cela peut entraîner un chargement plus long de la page.

5.2.2 Implémentation sur la page html talk

Les mêmes méthodes ont été utilisées pour implémenter la page talk avec les attributs de « Event ».

How to combine PROs and guidelines in your clinical practice: Illustrated with case studies. Ulcerative colitis



	Normal (%)	Malformation (%)	Pre-term (%)	Abortion (%)
Background population	83	2	6	9
Inactive CD	82 (71–93)	1 (0–6)	7	10 (3–27)
Active CD	54	1	25	20
Inactive UC	84 (76–97)	1 (0–3)	6	9 (1–16)
Active UC	65	2	12	21

1. Khosla R, et al. Gut 1984;25:52–6. 2. Nielsen OH, et al. Scand J Gastroenterol 1983;18:735–42. 3. Wilschey CP, et al. Gut 1980;21:469–74. 4. Miller JP, et al. R Soc Med 1986;79:231–6. 5. Dignass AU, et al. Dig Dis 2009;27:341–6. 6. Schreiber H, et al. Aliment Pharmacol Ther 2015;41:120. 7. van der Woude CJ, et al. J Crohns Colitis 2015;9:107–24.

Info Embed

Share this talk: [f](#) [t](#) [in](#) [m](#)

How to combine PROs and guidelines in your clinical practice: Illustrated with case studies. Ulcerative colitis

Axel Dignass, Agaplesion Markus Hospital, Frankfurt, Germany

vendredi, 18 mars 2016 · 12:51 après-midi · 01m 56s · 281 views · Amsterdam, The Netherlands



Figure 16 : Exemple de talk sur le portail de Klewel en serveur test (données de l'auteur)

5.2.2.1 Implémentation du RDFa

La problématique rencontrée dans l'implémentation dans la page conférence est aussi présente ici. Sans certaines connaissances, il peut être compliqué de retrouver facilement les balises concernées.

```
<div class="page-header" vocab="http://schema.org/" typeof="Event" style="...">
  <h1><span property="name">{{ talk.title }}</span><br/>
    <small>
      <span property="performer" typeof="Person"><span property="name">
        {{ talk.getPresentersWithSemicolons }}</span>, <span property="affiliation">{{ talk.getPresentersWithSemicolonsAffiliation }}</span></span>
      </small>
    </h1>
    <p class="lead">
      <span property="startDate" element="{{ talk.date_and_time|date:"Y-m-d" }}">{{ talk.date_and_time|date:"l, j F Y" }}</span>
      .
      {{ talk.date_and_time|date:"P" }}
      .
      {% with talk.getDurationAsTime as duration %}
        {% if talk.getDuration >= 3600 %}
          <span property="duration">{{ duration|time:"H" }}h
          {% endif %}
          {% if talk.getDuration >= 60 %}
            {{ duration|time:"i" }}m
            {% endif %}
            {{ duration|time:"s" }}s</span>
          {% endif %}
          {% with talk.getViews as views %}
            {% if views > 100 %}
              .
              {{ views|intcomma }} view{{ views|pluralize }}
              {% endif %}
            {% endif %}
          .
          <span property="location" typeof="Place"><span property="address" element="{{ webcast.location }}"></span></span>
        </p>
        {% if talk.summary %}
          <blockquote style="...">
            <span property="description">{{ talk.summary }}</span>
          </blockquote>
          {% endif %}
        </div>
```

Figure 17 : Section de la page HTML talk où le RDFa a été implémenté (données de l'auteur)

A la différence de la conférence, le talk ne possède pas de localisation propre vu que l'humain comprend implicitement qu'elle est identique. La machine n'a pas cette chance. La localisation du webcast a donc été repris et retransmis afin qu'un maximum d'informations puissent être fournies.

[illegible]

Figure 18 : Résultat du RDFa dans le code source de la page présentée en figure 16 (données de l'auteur)

Plusieurs éléments de schema.org n'acceptent des données que dans certains formats. C'est le cas par exemple pour les propriétés « `startDate` » ou « `duration` ».

Pour la date, le format défini est le suivant : Année-Mois-Jour en chiffre. Si ce n'est pas le format utilisé pour l'affichage, il est important d'insérer le format valide dans la balise via le terme « element ».

Pour le cas de la durée du talk, le problème dans l'affichage est l'ajout du terme « » qui est utilisé afin d'insérer un espace. Le format n'étant pas respecté, il faut aussi utiliser la même méthode que pour la date.

5.2.2.2 Implémentation JSON-LD

La plupart des éléments étant identiques à ceux de la page conférence, le script a pu être réutilisé et adapté au talk.

```
<script type="application/ld+json">
{
  "@context" : "http://schema.org",
  "@type" : "Event",
  "name" : "{ talk.title }",
  "description" : "{ talk.summary }",
  "startDate" : "{ talk.date_and_time|date:'Y-m-d' }",
  "duration" : "{% with talk.getDurationAsTime as duration %}{% if talk.getDuration >= 3600 %}
    {{ duration|time:'H' }}T{% endif %}{% if talk.getDuration >= 60 %}{{ duration|time:'i' }}M{% endif %}{{ duration|time:'s' }}S{% endwith %}",
  "performer" :
  {
    "@type" : "Person",
    "name" : "{ talk.getPresentersWithSemicolons }",
    "affiliation" :
    {
      "@type" : "Organization",
      "name" : "{ talk.getPresentersWithSemicolonsAffiliation }"
    }
  },
  "location":
  {
    "@type": "Place",
    "address" :
    {
      "@type": "PostalAddress",
      "name": "{ webcast.location }"
    }
  },
  "image":
  {
    "@type": "ImageObject",
    "url": "{ webcast.logo.link }"
  }
}
</script>
```

Figure 19 : Script JSON-LD implémenté dans la page HTML talk (données de l'auteur)

L'intégration de la propriété nom de la personne et celle du nom de l'organisation a été compliquée car Klewel avait développé une fonction qui retournait les deux informations en une seule chaîne de caractères. Nous avons donc dû créer deux fonctions afin que les informations soient séparées.

```

<script type="application/ld+json">
{
  "@context" : "http://schema.org",
  "@type" : "Event",
  "name" : "How to combine PROs and guidelines in your clinical practice: Illustrated with case studies. Ulcerative colitis",
  "description" : "",
  "startDate" : "2016-03-18",
  "duration" : "01M56S",
  "performer" :
  {
    "@type" : "Person",
    "name" : "Axel Dignass",
    "affiliation" :
    {
      "@type" : "Organization",
      "name" : "Agaplesion Markus Hospital, Frankfurt, Germany"
    }
  },
  "location":
  {
    "@type": "Place",
    "address" :
    {
      "@type": "PostalAddress",
      "name": "Amsterdam, The Netherlands"
    }
  },
  "image":
  {
    "@type": "ImageObject",
    "url": "http://www.ferring.com/"
  }
}
</script>

```

Figure 20 : : Résultat du script dans le code source de la page présentée en figure 16 (données de l'auteur)

5.2.3 Test

Afin de garantir une bonne implémentation, il est important d'utiliser des moyens de vérification afin de contrôler le bon fonctionnement du code.

5.2.3.1 Page Web webcast

L'affichage correct des informations que ce soit en RDFa ou JSON-LD ne garantit pas que les moteurs de recherche recevront correctement les informations. Il existe des technologies disponibles sur le Web afin de vérifier si tout est correct. La plus utilisée et performante actuellement est proposée par Google. Il s'agit d'un outil de test des données structurées (Outil de test des données structurées, s.d.). Il permet d'y insérer un morceau de code et il analyse automatiquement s'il existe des marquages et les différentes erreurs.

Tester le RDFa ou le JSON-LD devrait normalement nous fournir le même résultat si les deux méthodes ont utilisé les mêmes termes.

Nous avons premièrement testé le RDFa sur la page Web conférence.


EventSeries		0 ERREUR 4 AVERTISSEMENTS ^
@type	EventSeries	
name	Tailoring mild-to-moderate IBD management	
startDate	2016-03-18	
description	Patient-reported outcomes are becoming an increasingly important part of research and clinical management, but how can we integrate them with established clinical measures of disease activity? It gives us great pleasure to invite you to this Ferring-sponsored symposium. We shall explore mild-to-moderate Crohn's disease and ulcerative colitis case studies in the context of guidelines and clinical evidence as well as how patient-reported outcomes can help to align patient priorities with clinical practice, allowing you to tailor treatments to the needs of your patients.	
location		
@type	Place	
address		
@type	PostalAddress	
name	Amsterdam, The Netherlands	
 name	Il est préférable de remplir le champ <i>name</i> . Veuillez fournir une valeur si possible.	
about		
@type	Thing	
name	Health Sciences	
image		
@type	ImageObject	
url	https://search.google.com/structured-data/testing-tool/#__sid=rd3	

Figure 21 : Résultat du RDFa en figure 13 sur l'outil de Google. (Outil de test des données structurées, s.d.)

Nous pouvons constater qu'aucune erreur n'est présente ce qui confirme la bonne implémentation du RDFa.

Il était ensuite nécessaire de contrôler que les informations transmises avec JSON-LD étaient bien identiques.


EventSeries		0 ERREUR 4 AVERTISSEMENTS ^
@type	EventSeries	
name	Tailoring mild-to-moderate IBD management	
startDate	2016-03-18	
description	Patient-reported outcomes are becoming an increasingly important part of research and clinical management, but how can we integrate them with established clinical measures of disease activity? It gives us great pleasure to invite you to this Ferring-sponsored symposium. We shall explore mild-to-moderate Crohn's disease and ulcerative colitis case studies in the context of guidelines and clinical evidence as well as how patient-reported outcomes can help to align patient priorities with clinical practice, allowing you to tailor treatments to the needs of your patients.	
about		
@type	Thing	
name	Health Sciences	
location		
@type	Place	
address		
@type	PostalAddress	
name	Amsterdam, The Netherlands	
 name	Il est préférable de remplir le champ <i>name</i> . Veuillez fournir une valeur si possible.	
image		
@type	ImageObject	
url	http://www.ferring.com/	

Figure 22 : Résultat du JSON-LD en figure 15 sur l'outil de Google. (Outil de test des données structurées, s.d.)

Nous pouvons constater que les informations sont bien les mêmes ce qui confirme la bonne implémentation et les bonnes données seront transmises aux différents moteurs de recherche.

5.2.3.2 Page Web talk

Le même test a été effectué pour la page talk. Un premier contrôle fait avec le RDFa et un second avec JSON-LD.

Event	PRÉVISUALISER 0 ERREUR 5 AVERTISSEMENTS ^
@type	Event
name	How to combine PROs and guidelines in your clinical practice: Illustrated with case studies. Ulcerative colitis
startDate	2016-03-18
performer	
@type	Person
name	Axel Dignass
affiliation	
@type	Organization
name	Agaplesion Markus Hospital, Frankfurt, Germany
location	
@type	Place
address	
@type	PostalAddress
name	Amsterdam, The Netherlands

Figure 23 : Résultat du RDFa en figure 18 sur l'outil test de Google.

Comme pour la conférence, aucune erreur n'est à déplorer. Les informations seront correctement transmises (Outil de test des données structurées, s.d.).

Event	PRÉVISUALISER 0 ERREUR 4 AVERTISSEMENTS ^
@type	Event
name	How to combine PROs and guidelines in your clinical practice: Illustrated with case studies. Ulcerative colitis
startDate	2016-03-18
duration	01M56S
performer	
@type	Person
name	Axel Dignass
affiliation	
@type	Organization
name	Agaplesion Markus Hospital, Frankfurt, Germany
location	
@type	Place
address	
@type	PostalAddress
name	Amsterdam, The Netherlands

Figure 24 : Résultat du JSON-LD en figure 20 sur l'outil test de Google.

Le script JSON-LD propose exactement les mêmes informations. L'implémentation est donc correcte (Outil de test des données structurées, s.d.).

Un site ayant du Web sémantique doit normalement voir ses visites augmenter. Les recherches étant plus ciblées et avec plus de données, les moteurs de recherches devraient proposer des liens vers le portail de Klewel plus souvent. Cette analyse de la fréquentation n'est malheureusement pas possible durant le déroulement de ce travail étant donné que l'implémentation en production ne sera effectuée qu'après les échéances.

5.3 Sitemap.xml avec le framework Django

Comme expliqué, une partie de l'implémentation du sitemap avait déjà été réalisée par Samuel Dorsaz. La première partie de cette section va se focaliser sur l'explication de ce qui a déjà été fait et une seconde partie sur ce qui a été rajouté.

5.3.1 Sitemap des conférences

Lors de la réception de la machine virtuelle et du projet, l'implémentation nécessaire dans la classe conférence afin de les afficher dans le sitemap.xml était déjà réalisé.

La première étape pour afficher les éléments voulus dans un sitemap et d'intégrer une classe spéciale dans la classe concernée, ici conférence.

```
class ConferenceSitemap(Sitemap):  
    changefreq = "never"  
    priority = 0.5  
  
    def items(self):  
        return Conference.objects.filter(nice_url__isnull=False)  
  
    def location(self, item):  
        return reverse('tube_webcast', kwargs={'webcast_shortcode': item.nice_url})  
  
    def lastmod(self, item):  
        return item.publication_date_and_time
```

Figure 25 : Extrait du code implémenter dans la classe Conférence (données de l'auteur)

Plusieurs informations seront transmises grâce à cette classe. L'élément « changefreq » permet d'indiquer à quelle fréquence cet élément du sitemap va être modifier. Il est possible que les modifications soient journalières, hebdomadaires, mensuelles, annuelles ou comme dans l'exemple jamais.

Le terme « priority » indique l'importance de cette page par rapport à d'autres pages du site. Par exemple, la page principale du portail aurait une priorité de « 1 » étant donné que sans elle, il n'est pas possible de naviguer sur le reste du site. Les conférences ont été

définies avec une priorité de « 0.5 » car en supprimer une n'affectera pas les autres conférences mais effacera les talks intégrés. Pour cette raison, les talks auront une priorité de « 0.4 ».

La seule variable obligatoire dans cette classe est « items ». Elle doit renvoyer la liste des objets qui seront affichés sur le sitemap. Dans ce cas-là, les conférences possédant un « nice_url ». Cette propriété précise si la conférence est privée ou publique. Le sitemap ne devant afficher que les conférences publiques, nous fournissons à l'élément « items » uniquement ces webcast possédants un « nice_url ».

Le sitemap référence les URLs des éléments affichés, pour ce faire, il existe deux méthodes. La première, obligatoire, requière l'intégration de la méthode « get_absolute_url ». Celle-ci doit retourner l'URL qui s'affiche sur le navigateur Web.

```
def get_absolute_url(self):
    if self.nice_url:
        absolute_base_url = get_absolute_nice_base_url()
        if getattr(settings, 'NICE_URL_ABSOLUTE_PATH', None):
            # This is /watch/nice_url/conference_nice_url/
            relative_url = reverse('conference_nice_url', args=(self.nice_url,))
            # This is /watch/nice_url/
            nice_url_relative_url = relative_url.replace(self.nice_url, '')[:-1]
            return urlparse.urljoin(absolute_base_url, relative_url.replace(nice_url_relative_url, '/'))
        else:
            return urlparse.urljoin(absolute_base_url, reverse('conference_nice_url', args=(self.nice_url,)))
    else:
        absolute_base_url = get_server_base_url()
        return urlparse.urljoin(absolute_base_url,
                                reverse('tube_webcast', kwargs={'webcast_shortcode': self.shortuuid}))
```

Figure 26 : Fonction permettant de définir l'URL de la conférence (données de l'auteur)

Nous pouvons voir dans la figure ci-dessus que l'url retourné n'est pas le même si l'élément « nice_url » existe ou non.

L'autre méthode est d'ajouter un élément « location » directement dans la classe et d'y préciser l'URL à renvoyer.

Le terme « lastmod » permet d'indiquer la date de la dernière modification de l'item.

Une fois l'implémentation de la classe terminée, il faut intégrer le chemin vers le sitemap.xml dans l'application.

```

urlpatterns = patterns('',
    # Admin docs and admin
    (r'^admin/doc/', include('django.contrib.admindocs.urls')),
    (r'^admin/', include(admin.site.urls)),

    # Django piston API (for upload, etc)
    (r'^api/', include(settings.API_URLS)),

    # KlevelTube, KlevelCast, KlevelHub, whatever the name
    (r'^watch/', include('kleweltube.urls')),
    # (r'^watch_native/', include('kleweltube.urls')),

    # Widgets (for embed)
    (r'^widgets/', include('widgets.urls')),

    # (r'^watch_native/', include('kleweltube.urls')),
    (r'^$', lambda r: HttpResponseRedirect('watch/webcasts/')),

    # Projects
    (r'^projects/', include('projects.urls')),

    # Sitemap
    (r'^sitemap\.xml$', 'django.contrib.sitemaps.views.sitemap', {'sitemaps': {
        'webcast': ConferenceSitemap,
    }})
)

```

Figure 27 : Liste des redirections selon la zone du site voulu (données de l'auteur)

Les deux dernières lignes présentes sur la figure ci-dessus permettent d'accéder au sitemap via l'url : portal.klevel.com/sitemap.xml.

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" encoding="UTF-8" ?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>
      https://portal.klewel.com/watch/nice_url/01-10-2013-journee-valaisane-du-tourisme/
    </loc>
    <changefreq>never</changefreq>
    <priority>0.5</priority>
  </url>
  <url>
    <loc>
      https://portal.klewel.com/watch/nice_url/02-09-2013-mni-leipzig-complete-program/
    </loc>
    <changefreq>never</changefreq>
    <priority>0.5</priority>
  </url>
  <url>
    <loc>
      https://portal.klewel.com/watch/nice_url/03-06-2013-sspnet-summer-school-2013-complete/
    </loc>
    <changefreq>never</changefreq>
    <priority>0.5</priority>
  </url>
  <url>
    <loc>
      https://portal.klewel.com/watch/nice_url/03-09-2013-idiap-jain/
    </loc>
    <changefreq>never</changefreq>
    <priority>0.5</priority>
  </url>
</urlset>
```

Figure 28 : Résultat du sitemap.xml des conférences (Klewel, s.d.)

Le sitemap.xml liste chaque conférence publique présente sur le portail. Les informations de fréquence de changements ainsi que la priorité sont référencées pour chaque conférence.

Toute l'implémentation du sitemap pour les conférences était déjà intégrée dans l'application au début de ce travail.

5.3.2 Sitemap des talks

Dès les premières discussions avec Klewel, il était clair que réaliser un sitemap des talks allait être complexe. La structure de leur base de données posait problème quant au fonctionnement des différents packages sitemap proposés pour Django.

Pour qu'un sitemap fonctionne, chaque talk devrait retourner un URL unique et qui ne change jamais. Malheureusement, comme expliqué dans la section 7.1, un même talk peut être utilisé dans plusieurs sessions et donc posséder plusieurs URLs.

Pour comprendre, décomposons la structure des URLs du portail de Klewel pour les talks. Contrairement à la conférence qui affiche son nom dans l'URL, le talk affiche sa position dans la session concernée.

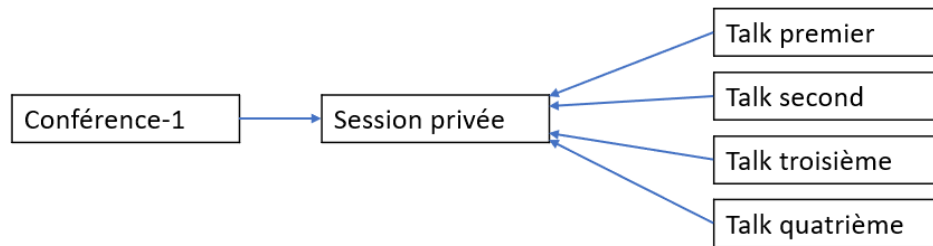


Figure 29 : Exemple graphique d'une session privée (données de l'auteur)

Dans la figure ci-dessus, l'URL pour le talk un serait :

<https://portal.klewel.com/watch/webcast/conférence-1/talk/1>

Figure 30 : Exemple d'URL d'un talk (données de l'auteur)

Le chiffre un en fin d'URL définit la position du talk dans la session. Il est donc possible qu'un même talk ait plusieurs positions selon la session à laquelle il appartient.

Prenons le scénario suivant : La conférence-1 possède deux sessions. La première, comme présentée dans la figure ci-dessus, est privée. Elle possède tous les talks. La deuxième session, imagée dans ci-dessous, est publique. Le client donne l'accès de tout utilisateur qu'aux troisième et quatrième talks.



Figure 31 : Exemple graphique d'une session publique (données de l'auteur)

Pour cette session publique, l'URL du talk troisième serait la même que la figure ci-dessus. Le talk étant en première position dans la session, le chiffre dans l'URL serait aussi le un. Nous pouvons donc constater que pour une même URL, deux talks peuvent être affichés. Le développement d'un sitemap pour les talks est donc impossible avec cette structure de base de données.

Conclusion

Bilan du travail

Plusieurs étapes furent nécessaires à la réalisation de ce travail de Bachelor. Il a d'abord fallu comprendre plusieurs notions importantes dans l'état de l'art. Une fois terminée, il a été possible de choisir une méthode de travail ainsi que les outils et technologies nécessaires.

Conséquemment à ces étapes, la partie d'intégration du web sémantique est concluante. L'objectif d'implémentation a été atteint. Les différents défis et obstacles rencontrés ont été résolus.

Toutefois, intégrer un sitemap référençant toutes les conférences et tous les talks n'est malheureusement que partiellement atteint. Seul la liste des conférences s'affiche. Aucune solution viable n'a été trouvée afin de contourner la problématique imposée par la structure de la base de données quant à l'affichage des talks.

Idées d'amélioration et perspectives

Une amélioration est évidemment de trouver une solution pour afficher les talks dans le sitemap. La première possibilité serait d'adapter la base de données afin que l'affichage soit possible. Une autre solution serait de trouver un moyen de contourner le problème.

Comme perspective d'avenir, Klewel aimerait travailler avec le site DBPedia afin de retourner des mots clés de chaque talk aux moteurs de recherche. Cette fonctionnalité, une fois existante, pourrait être implémentée dans le RDFa ou le JSON-LD.

Références

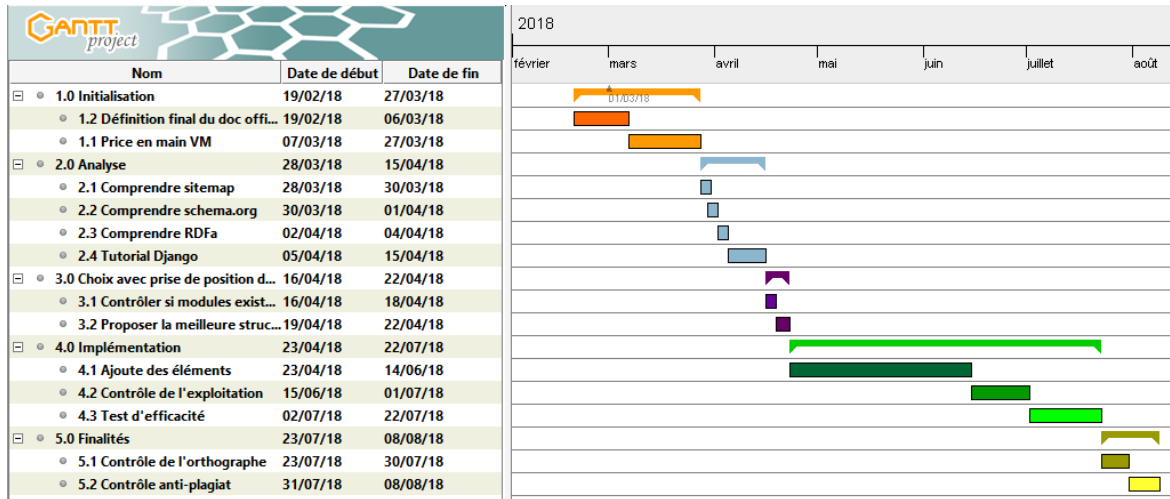
- Adida, B., & Birbeck, M. (2008, octobre 14). *Introduction à RDFa*. Récupéré sur yoyodesign.org: <http://www.yoyodesign.org/doc/w3c/xhtml-rdfa-primer/>
- Bathelot, B. (2015, août 2). *Définition : RDFa*. Récupéré sur definitions-marketing.com: <https://www.definitions-marketing.com/definition/rdfa/>
- Bathelot, B. (2017, mai 29). *Définition : Sitemap*. Récupéré sur definitions-marketing.com: <https://www.definitions-marketing.com/definition/sitemap/>
- Briche, C. (2018, janvier 12). *Comprendre les micros data et leur impact sur le SEO*. Récupéré sur seoh.fr: <https://www.seoh.fr/referencement-seo/comprendre-micros-data-impact-seo.html>
- Comprendre le web sémantique*. (s.d.). Récupéré sur commentcamarche.com: <https://www.commentcamarche.com/faq/14412-comprendre-le-web-semantique>
- Cuvelier, T. (2011, avril 6). *Web sémantique : introduction au RDFa*. Récupéré sur developpez.com: <https://tcuvelier.developpez.com/tutoriels/web-semantique/rdfa/introduction/>
- Event*. (s.d.). Récupéré sur Schema.org: <https://schema.org/Event>
- Guillemot, M. (s.d.). *Trends from Silicon Valley and learnings from a six times entrepreneur*. Récupéré sur Klewel.com: <http://www.klewel.com/blog/2016/09/22/trends-silicon-valley-loic-lemeur-laurent-haug/>
- HTML (HyperText Markup Langage) : définition, traduction*. (2018, juillet 13). Récupéré sur journaldunet.fr: <https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1203255-html-hypertext-markup-langage-definition-traduction/>
- Klewel*. (s.d.). Récupéré sur Klewel.com: <http://www.klewel.com/>
- Maignien, Y. (2014). *Chapitre 5. Les enjeux du web sémantique*. Récupéré sur books.openedition.org: <https://books.openedition.org/pum/320>
- Outil de test des données structurées*. (s.d.). Récupéré sur search.google.com: <https://search.google.com/structured-data/testing-tool?hl=fr>
- Person*. (s.d.). Récupéré sur Schema.org: <https://schema.org/Person>

- Quelle est la différence entre URI, URL et URN ?* (2018, juillet 15). Récupéré sur journaldunet.fr: <https://www.journaldunet.fr/web-tech/developpement/1203023-quelle-est-la-difference-entre-uri-url-et-urn/>
- Rouse, M. (2005, septembre). *Resource Description Framework (RDF)*. Récupéré sur searchmicroservices.techtarget.com: <https://searchmicroservices.techtarget.com/definition/Resource-Description-Framework-RDF>
- Schema.org : définition.* (2018, juillet 12). Récupéré sur journaldunet.fr: <https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1203519-schema-org-definition/>
- Shanon, V. (2006, mai 23). *A 'more revolutionary' Web*. Récupéré sur nytimes.com: <https://www.nytimes.com/2006/05/23/technology/23iht-web.html>
- Vallette, C. (2015, février 26). *JSON-LD : LA RÉVOLUTION SÉMANTIQUE DES MICRODONNÉES*. Récupéré sur woptimo.com: <https://woptimo.com/blog/2015/json-ld-la-revolution-semantique-des-microdonnees/>
- Web sémantique, web de données : définitions.* (2018, mars 15). Récupéré sur bnf.fr: http://www.bnf.fr/fr/professionnels/anx_web_donnees/a.web_donnees_definitions.htm
- I

Annexes

I. Diagramme de Gantt

Le logiciel GanttProject a été utilisé pour créer la planification de ce travail. La planification prévue a été respectée dans son ensemble.



II. Contenu du CD

Le CD ci-joint contient tous les fichiers utilisés durant le travail.

À la racine, vous trouverez le rapport en PDF (TB_Jonathan_Joaquim_2018) et le poster présentant ce travail de Bachelor.

Dans le dossier docs se trouvent :

- le document Word présentant les objectifs de ce travail (SOM_Klewel_JoaquimJonathan)
- Le diagramme de Gantt en format PDF (Planification)
- Le diagramme de classe complet de l'entreprise Klewel en format PDF (Class_Diagram_Klewel)

Dans le dossier projet se trouvent :

- Le projet réalisé (klewel-portal-TB)

Déclaration de l'auteur

Je déclare par ce document que j'ai effectué le travail de Bachelor ci-annexé seul, sans autre aide que celles dûment signalées dans les références et que je n'ai utilisé que les sources expressément mentionnées. Je ne donnerai aucune copie de ce rapport à un tiers sans l'autorisation conjointe du responsable de la filière et du professeur chargé du suivi du travail de Bachelor, y compris au partenaire de recherche appliquée avec lequel j'ai collaboré, à l'exception des personnes qui m'ont fournies les principales informations nécessaires à la rédaction de ce travail et que je cite dans la partie « Remerciements » de ce document.

Martigny, le 02 août 2018

Jonathan Joaquim