

Garagethon BlueHats des 9 et 10 janvier 2024 : « Codons ! »

Sommaire

En bref.....	2
Déroulé des journées.....	3
Projets représentés.....	5
Chatbot REL.....	6
Cunningham.....	6
ePoc.....	7
flop! EDT.....	7
FORDJ.....	8
GTNUM Forges.....	8
Issues.gouv.fr – New.....	9
Kubic.....	9
NDD Secteur Public.....	10
ONYXIA.....	10
Ontology pour la Valorisation des Données Publiques.....	11
PAD Pretty Printer.....	11
Ralph.....	12
Warren.....	12
Site Facile.....	13
Compte-rendus des sessions.....	14
Synthèse : TLDR ;.....	14
Mardi 14h - 15h Découverte Onyxia.....	15
Mardi 14h - 15h Usages de l'IA.....	15
Mardi 14h - 15h Présentation site.facile / Django - DSFR.....	20
Mardi 14h - 16h Dockerisation d'un projet django.....	20
Mardi 15h - 16h NDD Secteur Public listes noirs Univ Toulouse.....	21
Mardi 14h30 - 15h ePoc.....	22
Mardi 15h30 - 16h Présentation cunningham.....	22
Mardi 15h30 - 16h Atelier ontologie.....	22
Mardi 16h - Présentation de GTNUM - FORGES.....	22
Mercredi 10h30 - 11h30 Partage de bonnes pratiques pour initier une communauté de développeurs.....	23
Mercredi 10h30 - 11h30 Atelier Kubernetes.....	25
Mercredi 11h30 - 12h Quick fix accessibilité web.....	27
Mercredi 12h -13h Démon flash.....	27
Mercredi 14h - 14h30 Partage ressources éducatives libres.....	29
Mercredi 15h Informal Forge 15h 30.....	29
Mercredi 10h30 - 12h30 Où et comment trouver des sources de financement, Ospo?.....	30
Mercredi 14h - 15h code.gouv.fr Awesome : les projets à connaître.....	32
Mercredi 15h - 15h30 Atelier documentation pour code.gouv.fr.....	32
Mercredi 15h - 16h Atelier Wikidata.....	32
Bilan et prospective.....	33

Ce document est le compte-rendu détaillé du garagethon #BlueHats qui s'est déroulé les **9 et 10 janvier 2024** dans les locaux du [Lieu de la Transformation Publique](#), 77 avenue de Ségur à Paris. Ces deux journées étaient organisées par [la Dinum](#) et animées par [La Dérivation](#) et [l'Établi Numérique](#).

Ces [deux journées](#) s'adressaient aux **agents publics qui utilisent et contribuent à des logiciels libres** dans l'administration, membres de la communauté BlueHats ou non.

Elles faisaient suite à la [journée BlueHats "Mutualisons!"](#) du 1er juillet 2022.

En bref

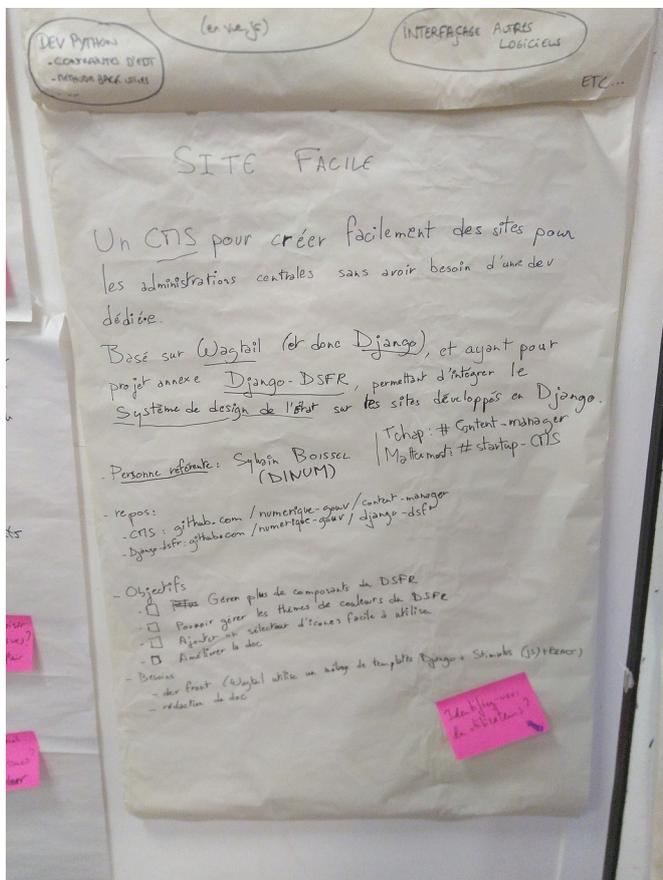
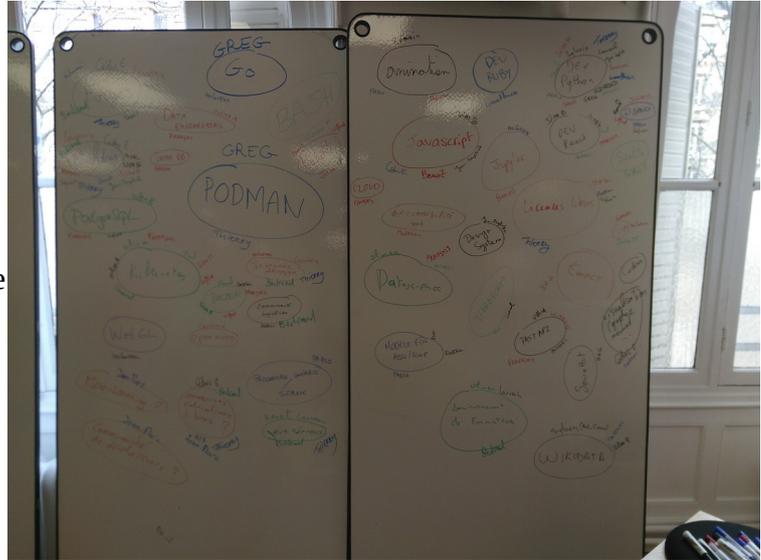
- Jusqu'à 25 personnes présentes en même temps, avec près d'**une quarantaine de participants sur les deux jours**
- Plus d'**une dizaine de structures** différentes représentées
- **18 projets** présentés
- **16 sessions** de travail, de 30 minutes à deux heures
- Des **thématiques abordées variées** ; petit échantillon : IA, ontologies et valorisation des données publiques, financement du logiciel libre, déploiement Django ou encore accessibilité web
- 7h de travail individuel ou collectif sur les deux journées
- **6 démos flash**
- Une **dizaine d'issues** créées et/ou résolues
- **Un prototype développé** pendant les deux jours (Pad Pretty Printer)
- Une quinzaine de chapeaux bleus de différentes formes distribués aux participants, ainsi que des exemplaires de la BD [Ada & Zangemann](#)



Déroulé des journées

Après un temps d'accueil et une présentation du garagethon et du contexte dans lequel il a lieu, la journée du 9 a commencé par un **speed meeting**, où les participants se sont déplacés dans l'espace pour aller à la rencontre des autres, pour des discussions de trois minutes maximum, en binôme.

Les participants ont ensuite été invités à remplir un grand **tableau de compétences**, inspiré par la méthode de la **Bulloterie**. Chaque personne pouvait ainsi choisir les capacités qu'il souhaitait mettre en partage sur les deux jours. Le tableau final a donné une préfiguration de la diversité des thématiques qui ont ensuite été abordées dans le reste du garagethon : de **FastAPI** au **modèle économique des assos et des coopératives**, en passant par les **jeux sérieux** et **Kubernetes**.



Les participants ont ensuite rédigé des panneaux de présentation des différents projets sur lesquels ils sont impliqués, à partir d'un canevas proposé par les animateur·ices. Ces panneaux sont retranscrits dans le chapitre *Projets représentés* de ce compte-rendu. Les participants ont ensuite pu prendre le temps de déambuler à la découverte de ces différents projets, avec la possibilité de poser des questions aux référents, par l'intermédiaire de post-its déposés sur les panneaux.

Tout au long de l'événement, les participants ont ajouté de nouveaux panneaux et mis à jour les panneaux existants, au fur et à mesure des arrivées et des travaux.

Ces premiers moments du garagethon avaient pour objectif de faciliter la rencontre et de susciter la curiosité des participants, afin de permettre de réaliser un programme collectif de travail, avec une méthode inspirée des [Forums Ouverts](#).

Les trois demies-journées du 9 après-midi, 10 matin et 10 après-midi ont ainsi été divisées en créneaux d'une demie-heure. Sur chaque créneau, il était possible d'organiser jusqu'à six ateliers simultanés dans des espaces de travail créés pour l'occasion et nommés selon des informaticiennes célèbres : [Chelsea Manning](#), [Susy Thunder](#), [Margaret Hamilton](#), [Grace Hopper](#), [Timnit Gebru](#) et [Ada Lovelace](#).

Une fois l'espace et le temps divisés, **chaque participant a pu proposer une ou plusieurs session de travail sur les créneaux disponibles**. Au fur et à mesure des propositions (et des ajustements éventuels), un programme complet a été construit, à partir des envies et des possibilités de l'ensemble des participants.



Projets représentés

18 projets étaient présents lors de cette session de travail, avec un panneau de présentation pour chaque projet.

Chatbot REL

Objectif

Créer un chatbot qui permet de faire des recherches sur des ressources éducatives libres

Descriptif

Un préalable : des outils pour aider des communautés de formation à répertorier et permettre l'exploitation de REL.
(Crawling, transformation dans un format ouvert type Markdown, indexation assistée par l'IA)

Front : des outils pour trouver et exploiter ces REL (recherche assistée par l'IA, réponse en langage naturel, recommandations, interrogation/analyse de la ressource, transformation en un autre format)

Mots-clés

#REL #IA #Markdown

Repo

Ø mais expérimentations autour d'un projet plus simple → ChatMD

Contacts

Cédric (DRANE Lyon)

Questions des participants

Statut de ChatMD ?

Cunningham

Objectifs

Disposer d'un Design System commun afin d'accélérer les développements, homogénéiser nos applications et avoir des interfaces accessibles.

Descriptif

Design System de FUN open source et customisable

Bien qu'à la base créé pour les besoins UI de FUN, sa fonctionnalité de customisation des designs tokens permet de customiser visuellement les composants pour l'adapter simplement à une autre charte graphique.

Les headless components nous permettent d'assurer une bonne accessibilité de nos composants.

Mots-clés

Atomic Design, Thème, Customisation, UI, Accessibilité, Headless components, Design Token

Liens

<https://github.com/openFun/cunningham>

Contacts

Jean-Baptiste, Antoine L., Manuel

Besoins / prochaines étapes

Proposition et développement de nouveaux composants.

Audit d'accessibilité et correctifs

Ouverture à d'autres framework js que React

Question des participants

Lien avec DSFR ?

ePoc

Descriptif

Application de mobile learning et éditeur de contenus Libres et open source. Disponible sur les stores.

→ Communauté de formateur

Contenus de types : texte, vidéo, audio, quiz, etc.

Accessible hors ligne

Mots-clés

Mobile learning, Ionic, Angular, Electron, iOS/Android

Liens

epoc.inria.fr

Contacts

Jean-Marc Hasenfratz

Benoit Rospars

ill-epoc-contact@inria.fr

On est sur Tchap : ePoc

Besoins / prochaines étapes

API Plugin à développer

Documentation technique

Création d'une communauté de développeurs

flop! EDT

Descriptif

Logiciel libre (AGPL v.3) de gestion coopérative des emplois du temps/planning.

Chacun définit ses contraintes et préférences

L'EDT est généré automatiquement (programme linéaire en nombres entiers)

Tout le monde peut modifier l'EDT (l'outil vérifie que les contraintes sont OK)

www.flopedt.org

Mots-clés

PostgreSQL

Django

D3.js → vue.js

CBC/Gurobi

Pulp

Liens

<https://framagit.org/flopedt/flopedt.git>

Contact

Pablo

Besoins / prochaines étapes

Recoder le frontend en vue.js (en cours !)

Solveurs opensource CBC/GLPK/Scip

Améliorer le déploiement / dockerisation

Sources de financement / Partenariats

Documentation / tutos

Dev. des interfaces manquantes (en vue.js)

Optimisation des requêtes BD / API-Rest

Dev Python

FORDJ

Descriptif

Une forge open source pour les architectes et les designers.

Fordj est une plateforme de conception (design d'espaces et de bâtiment) basé sur Forgego, un soft-fork de Gitea (un Gitlab écrit en Go)

Mots clés

Forge, 3D, Architecture, conception, BIMGIT, cherry-picking, rebase, Go, Infra / Hosting, e-learning, building information modeling

Liens

<https://fordj.org>

Contact

milovann (Écolmilovann), École d'archi Nancy /
Ministère de la Culture

GTNUM Forges

Descriptif

Surcouche d'une forge logicielle pour des enseignants, pour déposer et accéder à des jeux sérieux.

Mots clés

Jeux sérieux, Forges, Markdown, MKdocs

Liens

<https://lium.univ-lemans.fr/gtnum-forges/>

Contact

Thierry Forest

Besoins / prochaines étapes

Licences CCBy (NC)

Issues.gouv.fr – New

Objectifs

Mettre en avant des "good 1st issue"
Affiner des profils techs dans notre écosystème
Faire participer des citoyens, étudiants, candidats
Valoriser les contributions internes et externes

Mots clés

#Communauté
#Citoyenneté
#Gamification
#BlueHats

Contact

Julien B

Besoins / prochaines étapes

Feedbacks
UX
Issues de qualité sur les repos

Questions des participants

Comment cloisonner les dépôts issues ?
Sur quoi porte les issues ?

Kubic

Descriptif

Projet Terraform pour déployer du Kubernetes As A Service chez des fournisseurs de cloud souverains (OVH, Scaleway)

Objectifs

Transformer la solution existante en module terraform pouvant être importé dans des projets terraform déjà en place.
Rajouter d'autres cloud provider

Mots clés

#Terraform
#kubernetes
#argoCD
#helm

Liens

<https://github.com/openfun/kubic>

Contact

Manuel Raynaud

Question des participants

Besoins ?

NDD Secteur Public

Descriptif

Liste des noms de domaines (NDD)
d'organismes de secteur public.
Fin 2023 : 90000 entrées.

Objectifs

Intégrer la gestion de liste noire de l'université
de Toulouse → [MR en cours !](#)

Mots clés

#liste #reuse #SecteurPublic #Python

Liens

<https://gitlab.adullact.net/dinum/noms-de-domaine-organisme-secteur-public/>

Contact

Matthieu

Questions des participants

Besoins ?

Gouvernance ?

ONYXIA

Descriptif

"Datalab". Libre (à la demande) service
datascience : Jupyter (GPU), Rstudio (GPU),
VsCode (GPU), PostgreSQL

Liens

<https://datalab.sspcloud.fr> et
<https://github.com/inseefrlab>

Mots clés

Javascript (React)
Java (Spring boot)
Kubernetes (Helen)
Datascience (catalogues de services)

Contact

Olivier Levitt
Tchap #SSPCLOUD
innovation@insee.fr

Ontology pour la Valorisation des Données Publiques

Objectifs

Proposer aux administrations publiques un framework pour accéder la valorisation de leurs connaissances.

Description

L'abondance de données plus ou moins structurées au sein des administrations publiques complexifie leurs valorisation par les agents publics.

Des solutions existent pour structurer ces informations et les rendre actionnables : les ontologies, type Wikidata pour Wikipedia proposent des graphes de connaissances qui permettent d'exposer de façon standardisée de l'information contextualisée.

[SPARQL](#)
[Triplet Store](#)

Mots clés

Ontologie, Knowledge graph, NLP, Valo Data, Framework

Besoins / prochaines étapes

Feedback idée, Graphe, Wikidata, Retex

Question des participants

On part de quoi ?

PAD Pretty Printer

Objectifs

Provides a collection of well-designed templates for effortless printing of markdown PAD documents (Note PAD)

Description

How it works?

- 1 : copy/paste your markdown
- 2 : choose your template
- 3 : print your document

MVP

Support only markdown sources.
Choose among a list of existing templates.
Templates are shared with all (no user rights management).
Eternal service from pad.

Mots clés

Django (Python), NextJS (Typescript)

Contact

Sam Pacoud
Antoine Lebaud

Besoins / prochaines étapes

Just started, very WIP!

Ralph

Objectifs

Learning Record Store (LRS) qui permet de stocker des traces d'apprentissage (events émis par un apprenant lorsqu'il suit un cours)

Descriptif

C'est une HTTP REST API avec un principal endpoint `"/statements"`. POST `"/statements"` pour envoyer, GET pour fetch, etc.

Un LRS peut recevoir des statements (ou traces) d'un LMS (Learning Management System), ou d'autres composants.

Learning activity ex : classe virtuelle, composant vidéo...

Dataviz ex : Grafana (voir Potsie sous github.com/openfun), Warren (voir poster ci-après).

Backends compatibles : Elasticsearch, MongoDB, clickouse ou filesystem

Liens

<https://github.com/openfun/ralph>

La documentation est en cours

(openfun.github.io/ralph/dev/) pour la version 4.0

Besoins / prochaines étapes

Beaucoup de backends supportés par Ralph (ES, Mongo, clickouse, S3...), dur à maintenir !

Un retour et une prise de recul serait grandement apprécié :) (je paye ma bière :))

Warren

Objectifs

Learning analytics Framework

Descriptif

Calcul d'indicateurs

gestion de cache (indicateurs)

architecture plugins :

- * indicateurs

- * executive engines (Cronjob, RT, b&s CJ, AirFlow)

HTTP API to query indicators

React components (e-charts) dataviz

Datasources

- * LRS (learning Record Store) - xAPI

- * ?

Interopérabilité

LTI

Mots clés

Python, FastAPI, Dataviz, xAPI, Analytics

Liens

<https://github.com/openfun/warren>

Contact

Julien Maupetit (julien.maupetit@fun-mooc.fr)

Besoins / prochaines étapes

Extension des indicateurs disponibles

Exécution engines supportés

Support de différents datasources

Site Facile

Objectifs

Un CMS pour créer facilement des sites pour les administrations centrales sans avoir besoin d'une dev dédiée.

Descriptif

Basé sur Wagtail (et donc Django), et ayant pour projet annexe Django-DSFR, permettant d'intégrer le système de design de l'état sur les sites développés en Django.

Contact

Sylvain Boissel (DINUM)
Tchap : #Content-manager
Mattermost : #startup-CMS

Liens

CMS : github.com/numerique.gouv/content-manager

Django-dsfr :
github.com/numerique-gouv/django-dsfr

Besoins / prochaines étapes

Gérer plus de composants du DSFR
Pouvoir gérer les thèmes de couleurs du DSFR
Ajouter un sélecteur d'icônes faciles à utiliser
Améliorer la doc
dev front (wagtail utilise un mélange de templates Django + Stimulus (JS + React))

Question participants ?

Identifiez-vous des utilisateurs ?

Compte-rendus des sessions

Ces compte-rendus de session ont été réalisés par les participants de chaque session, avec un pad partagé mis à disposition de tous pendant les deux jours du garagethon. Merci à tous les participants pour cette prise de notes touffue et de qualité !

Synthèse : TLDR ;

Il n'est pas facile de résumer 16 sessions de travail aussi riches, mais on peut au moins tenter de faire le panorama des travaux.

Premier élément du panorama, les nombreuses sessions de présentation d'un projet ou d'un outil. Onyxia, Site Facile, NDD, ePoc, Cunningham, GTNUM Forges, Kubernetes et Wikidata : huit projets ont fait l'objet d'un temps de présentation détaillé. Ce garagethon a donc été un espace de partage de savoirs, où les participants ont découvert des pratiques et des outils nouveaux.

Par ailleurs, six projets qui ont fait l'objet d'une session spécifique (tous sauf Kubernetes et Wikidata) sont des projets développés par des membres de la communauté BlueHats. On peut ajouter à ça Sloyd, Pretty Pad Printer et FlopEDT, qui ont été présentés lors des flash demos du jeudi matin. La communauté BlueHats est donc bien une communauté de développement, qui porte au quotidien des projets libres utilisés dans des administrations. Par ailleurs, la complémentarité entre ces projets est souvent ressortie lors des discussions, avec des besoins futurs d'espaces nouveaux de coopération et de partage de problématiques.

Ce garagethon a justement permis des espaces de coopération, puisque plusieurs sessions de travail technique ont eu lieu : une demi-douzaine de participants ont travaillé pendant l'après-midi du mardi sur le projet floP EDT ! avec comme objectif de faciliter sa containerisation et donc son déploiement futur. Une session consacrée à la liste des noms de domaines du secteur public a débouché sur une issue et la production d'une merge request. Des ajouts et modifications ont été proposées pour le dépôt code.gouv.fr Awesome ainsi que pour le dépôt de document pour code.gouv.fr.

Un projet complet a aussi été lancé lors de ce garagethon : le projet Pretty Pad Printer, dont une version alpha a été présentée à l'issue des deux jours.

Ces deux jours ont aussi permis la tenue de discussion de fond au sein de la communauté :

- un échange de savoir et débat sur les usages possibles et pertinents de l'IA a eu lieu le mardi
- un partage de bonnes pratiques pour initier une communauté de développeurs a regroupé une demi-dizaine de personnes
- la question de savoir où et comment trouver des sources de financement pour les projets libres a occupé toute la matinée du mercredi

On peut remarquer aussi la place importante des problématiques d'éducation lors de ces deux jours : beaucoup de personnes travaillant dans ce secteur, et une bonne partie des projets présentés ont émergé directement de problématiques éducatives : gestion d'emploi du temps, e-learning, ressources éducatives libres, gestion de LMS, ...

Si on devait résumer, c'est la question de la communauté qui est le plus ressortie de ces deux jours, comme le bilan final l'atteste. Échanger des savoir et des outils, mutualiser des ressources et de

l'infrastructure par les forges, travailler sur des pistes communes de financement, bref pouvoir sortir la tête du silo dans lequel on peut se retrouver pour aller plus loin ensemble.

Mardi 14h - 15h Découverte Onyxia

Présentation [d'Onyxia](https://www.onyxia.sh/) - <https://www.onyxia.sh/>

Besoin de l'INSEE : "dématérialiser" le poste de travail des statisticien·nes / développeuses de l'INSEE. Sortir du modèle "je fais tourner mes calculs sur mon laptop" ; passer à "je faire tourner mes calculs sur une énorme machine, avec de conteneurs, de sorte que l'infra soit reproductible" et partageable (besoin similaire dans toute l'administration)

INSEE : ~1000 statisticien·nes (ça fait du monde)

SSP = Service de Statistique Public

Instance Onyxia de l'INSEE : <https://datalab.sspcloud.fr/>

Approche :

- les services sont volatiles
- le stockage est "permanent" (défini par les conditions de l'instance). Sur SSPcloud, c'est conservé mais sans garantie de service.

Permet de forcer la reproductibilité, et donc le passage en production.

Après clic-clic-clic, un fichier Helm est créé, et l'environnement lancé.

HElm = une espèce de `apt` pour Kubernetes

La hype qui claque du moment :

- DuckDB : permet de lire un format CSV ou Parquet (= CSV avec metadonnées)
 - permet de s'affranchir d'instancier un Postgresql pour "juste" requêter ses données
 - <https://duckdb.org/>
- Parquet (format de fichier) = CSV avec metadonnées. Utilisé entre autres sur Data.gouv.fr, format binaire orienté colonne idéal pour l'analytique.
- hype de la hype : DuckDB en WASM (web assembly)

SSPcloud :

- 22 machines
- 12 qui gèrent le front, chacune avec 100 CPU (!)
- plusieurs GPU (tout le temps utilisés)

Mardi 14h - 15h Usages de l'IA

Vous pouvez retrouver ici la synthèse de l'atelier sur les usages de l'IA, avec :

- le contexte de l'atelier
- les précisions de François Paupier sur trois manières de bien utiliser l'IA : le prompt engineering, le RAG, et le fine-tuning, ainsi que des recommandations sur des points techniques

- les notes brutes prises lors de cette session par Bastien

Contexte

Cédric Eyssette : chatbot REL

Projet de construction d'un **chatbot pour trouver et exploiter des REL** (Ressources Éducatives Libres) préalablement repérées et validées par une communauté de formation

Usages de l'IA envisagés dans ce cadre :

- **Crawling** de ressources assisté par l'IA pour identifier les éléments pertinents
- **Indexation** assistée par l'IA (repérage automatique de thématiques)
- **Recherche de ressources** en langage naturel (recherche sémantique : calcul de similarité), **réponse et recommandations en langage naturel**
- Possibilité **d'exploiter la ressource** (recherche sur la ressource) et de **la transformer** en un autre format (résumé, carte mentale, quiz ...)

Pour le moment, expérimentation autour d'un modèle plus simple (sans IA) d'un chatbot configurable en Markdown : <https://eyssette.github.io/chatMD>

Bastien Guerry : critères d'une IA libre

Questionnement dans le cadre de l'OSI (*Open Source Initiative*) sur les **critères de définition d'une IA libre** et donc sur l'usage d'une IA dans le cadre d'un logiciel libre

- **Conception maximaliste** : toute la chaîne de production d'une IA doit être libre (y compris les données d'entraînement)
- **Conception minimaliste** : il suffit que le fine-tuning et la réutilisation du modèle soient possibles

Bien utiliser l'IA (trois possibilités)

Synthèse sous forme de carte mentale.

[Voir schéma ici](#)

Prompt engineering

Définition

= réflexion sur la meilleure manière de formuler un prompt

Principe

Si on veut obtenir un résultat plus précis et bien adapté, une première technique est de bien concevoir son prompt, de savoir comment préciser sa demande et comment la mettre en place par étapes

Intérêt

- Ne demande pas de la ressource supplémentaire, mais simplement une réflexion et des tests pour déterminer les prompts les plus pertinents
- On peut facilement intégrer dans un outil qui fait appel à une IA générative des préprompts pour améliorer les recherches de l'utilisateur

Limites

Cela n'exclut pas le risque d'hallucinations (cas où l'IA 'a répondu, avec un ton confiant, quelque chose de faux). Or dans certains contextes (notamment institutionnels : Éducation nationale, Armée ...) on ne peut pas se permettre de diffuser des informations fausses.

Le *prompt engineering* n'est pas une garantie de vérité

RAG

Définition

RAG = Retrieval Augmented Generation (c'est de la recherche d'informations augmentée par l'IA)

Principe

On va utiliser l'IA seulement pour générer une réponse en langage naturel, mais on va la contraindre de générer cette réponse à partir d'un ensemble de documents qu'on lui fournit (pour avoir plus de garantie sur l'output)

Intérêt

- On définit une source de vérité afin d'éviter le risque d'hallucinations
- On peut circonscrire la recherche sur un périmètre défini et la cibler sur un corpus spécifique qu'on définit soi-même (contrôle de la pertinence des réponses)
- Mise en place qui reste accessible (voir ci-dessous : questions techniques), pas de besoin important en terme d'infrastructure, possibilité d'utiliser un modèle plus simple de LLM (l'intelligence est ici dans les documents, pas dans l'IA)

Points d'attention

- Il faut d'abord bien identifier la requête de l'utilisateur ("intent detection")
- Quand on constitue sa base de données, il faut bien savoir comment on découpe l'information et comment on va rendre possible le calcul de la similarité entre requête de l'utilisateur & documents pertinents (choix des modèles d'embeddings, calcul à la volée ou précalcul enregistré dans une base de donnée des vecteurs de mots des documents ...)
- Approche hybride possible : embeddings + recherche traditionnelle

Ressources

- Une ressource intéressante pour mieux comprendre le RAG : <https://gradient.ai/blog/rag-101-for-enterprise>
- Une IA qui fait en quelque sorte du RAG à partir de résultats internet : [Perplexity](#)

Une entreprise qui va développer ce type de technologie (projet à suivre) : Watura, dans le cadre de la formation aux métiers de l'eau, avec une subvention de l'État : pour développer l'usage de l'IA sur ce corpus très précis de formation.

Fine-tuning

Définition

Il s'agit d'affiner un LLM en ajoutant un corpus de données sur lequel on le fait travailler ([page wikipédia](#))

Principe

Entraîner *from scratch* un LLM est hors de portée (accessible seulement à de gros acteurs, demande énormément de ressources), tandis que le *fine-tuning* demande des ressources mais est envisageable

Intérêt

L'entraînement sur un corpus n'a pas pour but de limiter la recherche à des documents donnés et d'éviter le risque d'hallucinations, le *fine-tuning* est utile quand on veut que les réponses soient d'un certain type : niveau de langage particulier (connaissance d'un vocabulaire, distinction de niveaux de complexité dans la formulation ...), tâche particulière (génération d'un quiz avec un format spécifique, ...)

Point d'attention

- C'est plus coûteux en terme de ressources
- Cela suppose la constitution d'un corpus d'entraînement bien organisé et éventuellement l'utilisation d'un apprentissage par feedback humain (RLHF) ou via une IA (RLAIF)

Questions techniques

Ressources intéressantes pour comprendre le fonctionnement de l'IA

- <https://huggingface.co/blog/shivance/illustrated-llm-os>
- métaphore du [stochastic parrot](#)

Choix des modèles

llama2, mistral fonctionnent bien

Base de données

Base de données vectorielles : Qdrant, ChromaDB ; Milvus plus complexe.

On peut aussi utiliser une base de données plus classique et intégrer des vecteurs de mots.

Intégration de l'IA dans ses projets

Llama Index : pratique

LangChain : des abstractions trop lourdes et pas nécessaires

Notes de Bastien Guerry

Usages des IA génératives

projet : faire un chatbot de ressources éducatives

- questions/réponses en langage naturel
- réponses en fonction des ressources

Recherche par vecteur/similarité

RAG : recherche sur un doc

<https://gradient.ai/blog/rag-101-for-enterprise>

outils ?

base documentaire+IA generative -> chaine de traitement

RAG : Retrieval augmented generative

--> Pb des hallucinations

frein : IA généraliste -> comment les contraindre à ne parler que d'un sujet ?

Les enseignants donnent le corpus des contenus pertinents

Fine tuner ?

causal language modeling : apprend à trouver un mot manquant dans une phrase, selon les ressources qu'on lui a donné (ex: contenus donnés par un prof d'histoire)

[Stochastic parrot](#)

Approche RAG

- un moteur de recherche classique qui retrouve des paragraphes
- les paragraphes sont donnés au LLM
- --> synthèse faite par le LLM en fonction des ressources d'apprentissage

Approche hybride

Projet RAG : des centaines de logiciels libres

Sentence piece

IA open sources ? 4 libertés pour un système AI

Faut-il inclure les données d'entraînement ? des débats autour de cette question

2 approches

- embedding sur tous ses documents -> très lourd
- générer le embedding à la volée

Des exemples : ...

Pour le LLM, [Mistral](#) marche très bien

Usage:

- communauté d'enseignants
- en ciblant des domaines
- outil d'interrogation de ressources

Au niveau des armées (réflexion depuis nov 2022), niveau expérimental

[fine tuning](#)

llama index : framework qui facilite l'injection de données dans un RAG

[Perplexity](#) --> moteur de recherche AI (RAG, recherche internet)

LLM OS

Autres approches

- Prompt engineering
- passer de l'information au contexte
- fine tuning où on entraîne un modèle

[ChatMD](#) -> chatbot à partir de Markdown

Notes au fil des échanges

Cédric : quels usages de l'IA pour coder ? Que pourrait-on faire avec des modèles d'IA ?

On peut imaginer de l'indexation assisté par l'IA, du crawling assisté par l'IA pour identifier des éléments pertinents, etc.

On peut imaginer un ChatBot en langage naturel, et ça renverrait vers des propositions en langage naturel. Ça ne chercherait pas sur le Net mais sur la base d'entraînement. Comment faire de la recherche sur un document à partir de l'IA (rag)?

François : plein de choses à partager en best practice. Retrieval augmented generation : utiliser l'IA générative pour faire des résumés à partir d'une base de recherche limitée. Dans un cadre où il y a une forte aversion au risque, on doit lutter contre les hallucinations : le mécanisme du rag fait la jonction des deux mondes : IA générative et base de données limitées. On stocke et indexe la base documentaire d'une certaine manière pour que le modèle fasse la synthèse en se bornant à ce contexte-là.

Cédric : On a aussi besoin de réponses justes.

Gauthier : dans les projets auxquels j'ai été confondu, l'un des problèmes principal était le coût du détournement : comment on contraint une IA à ne parler que d'un sujet particulier.

Jean-Marc : il y a une startup qui s'est présentée à EducaTech [Alexis Kauffmann se souviendra peut-être du nom]: les professeurs avaient référencé des ressources de vérité, l'interaction se faisant dans une vidéo générée par l'IA mais avec un périmètre délimité.

Cédric : Qu'est-ce qui est le plus pertinent entre fine-tuner une IA sur un corpus et faire du rag ?

François : il y a une subtilité : quand tu fine-tunes un modèle, par exemple avec des documents historiques choisis par des professeurs d'histoire. Le modèle va prendre le texte, cache des mots et essaie de deviner les mots cachés : donc là si tu fine-tunes sur une thématique spécifique, tu n'as pas de garantie de véracité, on a juste une suite probable de mots. C'est intéressant si tu as besoin d'un modèle pour une audience ayant besoin de lire dans un style précis. [Voir le papier de Timnit Gebru sur le "stochastic parrot"]

Bastien : est-ce que l'approche rag permettrait d'exclure des résultats de Copilot les propositions de code qui sont sous telle ou telle licence ? [Pas évident.] Est-ce que des acteurs se positionnent dans le domaine pédagogique ?

Cédric : il y a un avis CMA autour des ressources libres et de l'IA, dans ce cadre là on discute, on a rencontré une entreprise à Lyon qui commence à faire de la recherche sémantique sur des ressources web. L'un des problèmes de ces sociétés, comme elles dépendent toutes d'OpenAI et de ChatGPT c'est qu'elles n'ont pas la possibilité d'évoluer. Dans ce qu'on imagine, on aimerait se greffer sur un LLM open source européen (comme Mistral, par exemple).

Bastien : Difficulté de la définition « d'Open Source AI. » On doit pouvoir [utiliser/étudier/modifier/partager] tout le système, [y compris les données d'entraînement, pour la version maximaliste]. Certains suggèrent que la possibilité du *fine-tuning* suffit pour considérer qu'on "modifie" un modèle.

Cédric : questionnement sur comment toucher des communautés de formation en ciblant des domaines ? Un des domaines ciblés peut par exemple être celui des étudiants en santé. Est-ce qu'il y a déjà des usages côté Armées ?

Mardi 14h - 15h Présentation site.facile / Django - DSFR

Objectif de Site Facile : permettre de créer des sites pour des services sans avoir besoin de développeurs, sans faire un appel d'offres à un sous-traitant.

Solution basée sur Wagtail, un CMS en Django et avec un gros focus sur l'accessibilité.

Liens :

- Site de présentation : <https://content-manager.osc-fr1.scalingo.io/>
- Site d'exemple d'une installation neuve : <https://content-manager-staging.osc-fr1.scalingo.io/>
- Repo : <https://github.com/numerique-gouv/content-manager>
- Site de Wagtail : <https://wagtail.org/>

Objectif de Django-DSFR : permettre d'implémenter facilement le système de design de l'État sur un site utilisant Django

- Repo : <https://github.com/numerique-gouv/django-dsfr>
- Documentation : <https://numerique-gouv.github.io/django-dsfr/>

Mardi 14h - 16h Dockerisation d'un projet django

Repo sur <https://framagit.org/flopedt/FLOpedT>

Découverte des deux docker compose de flop!EDT : production et développement
Le docker compose de développement a l'air fonctionnel, sauf pour le service vue.

Note : Une bonne chose serait d'ajouter un wrappeur script bash `compose` pour wrapper l'exécution du docker compose, avec set des variables d'environnements, puis exécution du docker compose avec les arguments (avec "\$@").

Note : il serait bien de renommer le fichier `docker-compose.development.yml` en `docker-compose.yml` pour qu'il soit considéré par défaut, et d'éviter de le spécifier à chaque commande docker compose.

- Ajout d'une règle Makefile `bootstrap`
- Freeze des requirements/dépendances
- Ajout d'une entrée `help` dans le Makefile pour documenter chaque règle

Conseil : enlever le dockerfile Vue, utiliser une image node (lts, version 20.10), ajout d'une règle Makefile `install-frontend` pour `yarn install --frozen-lockfile`

Mardi 15h - 16h NDD Secteur Public listes noirs Univ Toulouse

<https://gitlab.adullact.net/dinum/noms-de-domaine-organismes-secteur-public>

On travaille sur l'issue [[#56 Utiliser le service de liste blanche / liste de noire de l'université de Toulouse](#)].

https://gitlab.adullact.net/dinum/noms-de-domaine-organismes-secteur-public/-/merge_requests/55

Au final seulement 4 matches après vérification de la liste noire, dont 2 sont des faux positifs. Ça vaut probablement pas le coup de mettre en place un filtrage régulier à partir de la liste de Toulouse.

```
import csv

publics = {}
with open("domains.csv", newline="") as csvfile:
    spamreader = csv.reader(csvfile, delimiter=",",
    quotechar="|")
    for row in spamreader:
        try:
            publics[row[0]] = True
        except:
            pass

with open("domains") as domainsfile:
    domains = domainsfile.readlines()
    for line in domains:
        if line.strip() in publics:
            print("MATCH", line.strip())
```

```
~/Projects/Bluehats ...
[I]→ python main.py
MATCH bestfreecams.club
MATCH blauzac.com
MATCH saintmarcel.com
MATCH sarrigne.fr
```

Mardi 14h30 - 15h ePoc

<https://epoc.inria.fr/>
mobile learning facile

Mardi 15h30 - 16h Présentation cunningham

Design system de France Université Numérique, customisable et accessible.

- Repo : <https://github.com/openfun/cunningham>
- Documentation : <https://openfun.github.io/cunningham>

Mardi 15h30 - 16h Atelier ontologie

- [RDF](#) pour la constitution pas pour la contribution
- [Wikidata](#) le site sur lesquels les données sont stockées
- [SPARQL](#) le langage de requête
- Projets Wikidata créé en Allemagne, bcp de contributeurs Fr, Allemands et néerlandais à ses débuts du projet
- Ex dans une page Wiki de [Leonardo Da Vinci](#), la partie droite sont un cartouche sur la partie droite
- Pour tout entrée dans Wikidata, on a soit une "nature de l'élément" (pour objet concret), "sous-classe de" (pour classe de l'arborescence)
- Pour la validation des propriétés (qui définit l'ontologie), il y a un vote au sein de la communauté Wikidata
- Supporte les références
- Avertissement dans la page Wikidata si il manque des données qui devraient est référencées
- On utilise des valeurs nulles quand on sait qu'il n'y pas valeur pour une relation;
- Un des points fort de Wikidata est de permettre de renseigner des infos de références sur une entité et de servir de pivot du web sémantique (alias/index)
- MediaWiki, extension [WikiBase](#) a prendre en compte pour instance une initiative pour créer une base de connaissance collaborative;
- [Moteur BlazeGraph](#) pour exécuter une requête (BDD Graphe).

Mardi 16h - Présentation de GTNUM - FORGES

Qu'est-ce qu'un GTnum ?

GTnum [Groupe Thématiques numériques](#)

Appel à projets de 3 ans du Ministère de l'Education Nationale et de la Jeunesse.

Communs numériques au service de la communauté éducative : modalités de production et de partage , organisation de la gouvernance, facteurs et modalités fcailitant la transformation des pratiques et le passage à l'échelle.

Qu'est-ce qu'une forge ?

Une forge est une plateforme en ligne de partage et de collaboration autour de documents textuels. Actuellement principalement utilisée par des informaticiens (le code, c'est du texte), elle pourrait être intéressante pour d'autres contenus (cours, diaporamas, exercices, corrections, tests, figures, ressources, logiciels, etc.), car elle permet de publier des contenus accessibles à tous, y compris ses différentes versions

Qui sont les acteurs du GTnum-Forges ?

- Laboratoires universitaires :
 - LIUM (Laboratoire Informatique de L'Université du Mans)
 - IUT de LAVAL

- CREN (Centre de Recherche d'éducation de Nantes)
 - Nantes - Angers - Le Mans - Laval
- CUREJ
- Institut national supérieur du professorat et de l'éducation (INSPÉ)
 - INSPÉ de Nantes
 - INSPÉ de Reims
- Rectorats
 - Pays de la loire
 - Bretagne
- Le réseau Canopé
- La mission Laïque Française

Organisation

Le GTnum est piloté par le LIUM <https://lium.univ-lemans.fr/en/gtnum-forges/>

1. Analyse des pratiques de partage et de collaboration des enseignants
2. Co-conception avec les fonctionnalités des forges
3. Evaluation des outils
4. Co-construction de formations

Mercredi 10h30 - 11h30 Partage de bonnes pratiques pour initier une communauté de développeurs

- DINUM
 - Content Manager -> Site Facile tech:Wagtail Django (Actuellement maintenu par Sylvain Boissel)
 - 1 an d'existence : Pas vraiment de communauté (Quelques utilisateurs motivés contribuent sur le projet.)
- INRIA
 - L'institut n'est pas une société de production de logiciel.
 - Comment maintenir une application développée dans un cadre de recherche ?
- LIUM Laboratoire de L'université du Mans
 - Production de Numérique éducatif (Laboratoire propriétaire par défaut)
 - Cycle de vie d'un produit de recherche
 - Hypothèse
 - prototype pour prouver l'hypothèse
 - Destruction du prototype
 - Comment pérenniser le produit des découvertes
- Le projet GTNum-forges à pour but
 - Etudier les usages des enseignants
 - Adapter ces usages à une interface
 - Expérimenter
 - Développement de communs que les enseignants puissent s'approprier ces communs
 - Développement d'une communauté d'acteurs (enseignants producteurs de contenus, développeurs de fonctionnalités)

Une communauté, une forge commune pour centraliser les productions.

Comment faire des communs ?

--> Les développeurs s'investissent lorsque le projet est pérennisé !

Pas de tests double aveugle possible :

Situation écologique

Etude qualitative car les résultats ne sont pas toujours reproductibles.

Objectif : ne pas démobiliser les acteurs.

Comment faire une communauté ?

SAT (Société d'intégration technologique)

Ces structures sont réticentes dès que le sujet des licences est abordé.

Ed Tech

Un Laboratoire Informatique n'est pas une société de production de logiciels.

Possibilité de pérenniser les productions avec des partenariats avec EdTEch.

Budgets de maintenance différents des budgets de réalisation. Plus difficile à définir.

Collaboration avec les EdTech : problématique de la propriété du code.

Fédérer les forges :

Problèmes d'accès aux forges

Propagation des Issues sur les forges

<https://forgejo.org/>

Activity pub 'fediverse'

<https://forgefed.org/>

Git lab

Que les Issues puissent être transmis à toutes les forges.

MOOC financé pour faciliter l'utilisation des forges

INRIA -

Solutions :

Embaucher des animateurs de communautés

Ces activités ne sont plus des activités de recherche

Fiche de poste de certains ingénieurs pédagogiques.

Producteur de contenus (fonction)

<https://www.animafac.net/> association d'étudiants ont les mêmes problématiques

[renater](#) serait-elle une structure possible pour fédérer les forges ?

- Modèle possible : Utilisation d'une association (<https://adullact.org/>)

- Pour assurer la maintenance des logiciels libres

- Possibilité d'utiliser EPOC pour l'adosser à une forge ?

- Problématique du stockage

Question des enseignants ? Comment gérer les vidéos?

Youtube, Peertube ?

<https://polymny.studio/> (scop)

- Ils hébergent les vidéos.

Quel est le modèle économique idéal pour un logiciel libre ?

- Firefox

-

Rôles:

Un animateur de communauté

rassembler une

relire les pairs

répondre dans le forum

triage des issues

compétences (com, gestion de projet)

Communiquer avec les utilisateur (community management)

Développeurs pour faire les taches non réalisés par les contributeurs.

Pistes de financements:

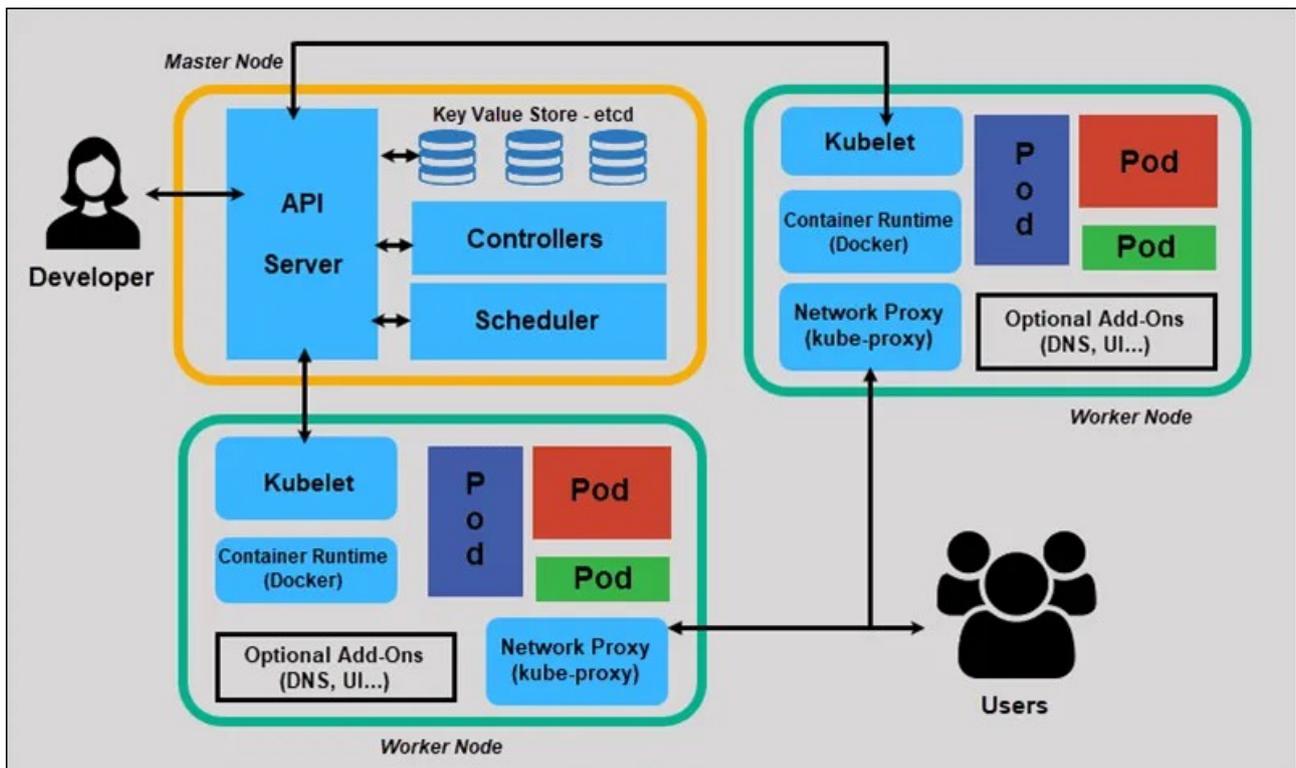
Dons d'entreprises utilisatrices

Institution publique ex: Big blue button (DNE)

Mercredi 10h30 - 11h30 Atelier Kubernetes

Généralités

Kubernetes est un orchestrateur pour des solutions de conteneurisation implémentant CRI ([Container Runtime Interface](#))



Description :

- 1 nœud = 1 VM ou une machine physique, avec des pods et un agent Kubernetes (Kubelet)
- 1 worker = 1 OS avec un *Container Engine* (docker, containerd, etc)
 - le *kubelet* est "l'agent infiltré" qui passe son temps à faire `docker ps` et faire un rapport au master
 - 1 pod = 1 conteneur (en première approche)
 - 1 worker peut être ajouté/supprimé **à chaud**
 - il y a entre **zero** et N workers
- 1 master (ou *control plane*)
 - API-Server : API standard avec laquelle le développeur interagit, avec laquelle ils interagissent avec les workers et récupèrent des informations des workers
 - le client de référence de l'API server est [kubectl](#)
 - Il y a entre **un** et N master (il y a forcément un chef). Il est recommandé d'en mettre un nombre impair: 1, 3 ou 5, souvent c'est 3. (culture confiture cf [Généraux byzanthins](#))
 - l'état du cluster est stocké dans du *etcd* (= entrepot clé/valeur)

Openstack est un orchestrateur de **VM**, mais pas du tout de conteneur. On peut cependant ajouter une orchestrateur de conteneur *par-dessus* un Openstack.

Distribution

<https://github.com/kubernetes/kubernetes>

Une distrib Kubernetes facilite la vie sur l'installation, les mises à jour et éventuellement ajoute des goodies.

Exemples de distrib:

- on premise
 - [Kubernetes the hard way](#), comme son nom l'indique...

- [Minikube](#)
 - la plus simple
 - en local sur une machine
- [k3s](#)
 - simple
 - débute avec un noeud local, peut supporter d'autres ajouts de noeuds
- [kubespray](#)
 - Production ready
 - basé sur Ansible (c'est juste des scripts ansible qui installent du Kube)
 - c'est la distrib utilisée par l'INSEE
 - c'est pas très opinionated
- [Openshift](#)
 - Distribution propriétaire redhat
 - Des ajouts en + du standard Kubernetes (client `oc` à la place de `kubectl` ...)
- cloud public
 - offres OVH
 - prenez vos VM cloud et installez-le vous-même (ce qui revient à le faire on premise)
 - offre Managed Kubernetes
 - limites : on n'a pas accès au master, donc il existe quelques configurations qu'on ne peut pas faire
 - avantage : simple à mettre en place, support de l'autoscaling
 - prix = prix des instance cloud (wouaw !). Google peut ajouter un markup sur GCP

Autoscaling :

- horizontal pod auto-scaler (géré par Kubernetes) ;
- node auto-scaler non géré par Kubernetes (intelligence fournie par les Cloud Providers), car pas de sa responsabilité

Mercredi 11h30 - 12h Quick fix accessibilité web

outil pour tester l'accessibilité : [dépôt Asqatasun](#)

Instance Adullact <https://asqatasun.adullact.org/>

Sites à tester :

- <https://content-manager-staging.osc-fr1.scalingo.io/>
- <https://datalab.sspcloud.fr/>

Mercredi 12h -13h Démonos flash

Présentation Pad Pretty Printer

Voir poster du projet

ANCT

[ANCT](#)

Grist, un airtable open source : <https://github.com/gristlabs/grist-core/> ; exemple d'un outil de centralisation national des ORT, où le système de permission très fin (à la cellule) permet de partager un document avec des possibilités d'éditations limitées ; possibilité de faire des vues customs pour l'utiliser dans des intégrations web

Contribution à un projet open source existant.

Version : Yunohost , Desktop

Problématique :

- Test de charge

Sloyd

sloyd.work, une plateforme d'apprentissage de la conception numérique : <https://sloyd.work/> ; basé sur Forgejo

- Ressources documentaires (wiki)

-

- Construire une ville : SIM City (100m/100m)

- Utilisation de blender, python, markdown

Système automatique d'évaluation:

- Utilisation d'un serveur Python pour évaluer les résultats des étudiants

- Utilisation du wiki pour le dépôt des travaux des étudiants./

- Affichage des résultats des étudiants

Génération automatique de l'intégralité des travaux des étudiants

ePOC JEan-Marc INRIA Appli+éditeur de mobile learning

- problématique : comment pérenniser, faire communauté

- Ressources (textes, vidéo) + quiz

- orienté objectifs péda

- système de badge avec contions (ET)

- C'est sur les stores

- C'est orienté vie privée

- L'éditeur a beaucoup de succès « il est ultra simple »

Gestion des badges : norme [xAPI](#) : j'ai vu une vidéo, j'ai un badge. J'ai répondu correctement correctement à un QCM, j'ai un badge.

FlopEDT

IUT de Blagnac

Centré auto-gestion (chacun peut faire son EDT, donner ses contraintes sans intermédiaire)

Flash présentation du GTnum Forges

<https://lium-cloud.univ-lemans.fr/index.php/s/49867zAcHan8ydc> (ce document est normalement destiné aux enseignants)

Observatoire de l'accessibilité du web

collecte des domaines du secteur pub → moulinette (qui détecte les erreurs grossière) → résultats

33000 pages d'accueil. En moyenne 34, en médiane 20.

1/3 des erreurs sont de contraste. Principalement des erreurs faciles à corriger.

Il faut agir. Il faut afficher le référent accessibilité.

Références :

- <https://adullact.org/breves/68-actualite/vie-associative/1103-l-observatoire-de-l-accessibilite>
- [Asqatasun](#)

Mercredi 14h - 14h30 Partage ressources éducatives libres

Charles POULMAIRE - Président AEIF

Vincent Xavier JUMEL

Milovann YANATCHKOV -

Matthieu GUERIN - DNE éduc

Bertand MARNE - LIUM

[capytale](#)

REL (Ressources éducatives libres) : [Définition \(en anglais\) sur Wikipédia](#) et [définition de l'UNESCO](#)

[Colin de la Higuera](#): les 5R (*traduction*)

- retenir
- réutiliser
- réviser
- remixer
- redistribuer

<https://fabriquerel.org/wp-content/uploads/2020/06/REL-5R-fabriqueREL.pdf>

Forum : <https://mooc-forums.inria.fr/moocnsi/login>

- Espace d'échange
- Structuré, filtré
- Ressources : Exercice

<https://forge.aeif.fr/e-nsi/pratique>

[Forge des communs numériques](#) de l'Education Nationale.

Ex. de REL : [Cours de NSI](#)

Cours de Maths : <https://coopmaths.fr/alea/>

Cours de python : <https://python.sdv.univ-paris-diderot.fr/>

Un logiciel peut être considéré comme une REL.

Prof d'histoire

Problématique avec les mineurs

- RGPD
- Production des ressources par les élèves

-

Etudiants disposent d'un dépôt

changer les habitudes des enseignants

Mercredi 15h Informal Forge 15h 30

Vincent-Xavier

Bertrand Marne
Bastien Guerri (GNU
Technique
Mailling-list : dépôt

Silo de forge universitaire
<https://sourcesup.renater.fr/>

Fo

Problèmes des Forges qui ne sont pas orientés pour l'édition.

Linux kernel Mailling list

ESR rapport sur les forge (150 pages) Comparatif fonctionnalités déploiement

Rapport de 50 Pages > <https://www.enssib.fr/bibliotheque-numerique/documents/71398-forges-de-l-enseignement-superieur-et-de-la-recherche.pdf>

Réseau social --
Ressources éducative libre

Forge

Approche différente
Former les futurs utilisateurs

Interface :
- Pratiques (outils)
- Partage des contenus et collaboration

Fallback
cout de maintenance github

Mercredi 10h30 - 12h30 Où et comment trouver des sources de financement, Ospo?

Pablo : Je me positionne dans le champ des éditeurs d'un logiciel libre. La construction d'un modèle économique n'est pas évident, quand tu ne factures pas des services. Il n'y a pas de clef et de modèle évident, tous ces modèles sont source d'effort. Est-ce que par la voix des institutions on peut trouver des financements ? Si oui, lesquels ? Est-ce par la communauté des éditeurs qui peut soutenir ?

Par exemple, nous [Flop!EDT](#), on a besoin de passer un seuil critique. Voilà mon point de départ avant de partager nos doutes et questions. Au début je pensais qu'il suffisait de faire un truc cool pour que ça marche. En tapant "logiciels libres financements" j'ai trouvé que la région Nouvelle-Aquitaine dédiait 300K€ pour ça et on a obtenu 40K€. Et sinon les utilisateurs qu'on accompagne paient 3K€. J'ai essayé de contacter plein de structures mais c'est très chronophage.

Milovann : Le libre est associé à l'amateurisme vs professionnalisme.

Bastien : NGI, Next Generation Internet met de l'argent pour soutenir des projets libres. NLnet est l'un des opérateurs qui gère les appels pour NGI. OW2 en est un autre.

Milovann : Il y a des entreprises qui affichent leur soutien.

Vincent : À l'ANCT, Agence Nationale de la Cohésion des Territoires, il y a un incubateur de services numériques ("Startup d'État") et nous avons une équipe pour accompagner les territoires autour de la donnée. Irritant identifié : tous les fichiers .xlsx dans les boucles de mail. On cherchait un tableur collaboratif, on a identifié NocoDB, Baserow et pour nous, on voulait partir d'un outil libre. Aujourd'hui, on a une équipe (deux développeurs et un designer) pour contribuer à [Grist](#). Nocodb, sous AGPL, est réticent à la contribution alors que nous on voulait pas forker. Notre seule réticence pour Grist c'était que leur business model était open core. L'équipe Grist était très ouverte à la collaboration et deux ans plus tard on contribue. On les a invité à [Numérique en Commun\(s\)](#) pour se connaître.

Bastien : BlueHats Semester of Code a permis de mobiliser des stagiaires pour des projets, et plusieurs nous ont dit que c'était super utile, car sans "surcoût" pour le projet libre qui reçoit. Donc distinguer les "aides" des "financements".

Pablo : Cette aide peut être piègeuse si on recrute le mauvais stagiaire.

Vincent : Mon besoin aujourd'hui est d'être capable de répondre à Grist qui se demande comment faire évoluer leur business model. Quels moyens a-t-on que des codes sources auxquels on contribue activement ne se referment pas ? Notre interlocuteur est ultra flexible.

Bastien : On pourrait avoir des acteurs privés qui soutiennent des projets libres (un station F du libre) mais on pourrait aussi avoir une "fondation BlueHats" qui aiderait les acteurs publics à bien interagir avec les entreprises pour s'assurer que celles-ci sont des acteurs vertueux dans le libre (c-à-d ne referment pas le code.)

Vincent : Notre contexte c'est celui des startups d'Etat où on expérimente, et on est encore là dedans : on a cette chance de pouvoir recruter des développeurs qui sont autorisés à contribuer à du libre, mais ce n'est pas généralisable.

Matthieu : Pour rebondir sur les modes de financement qu'on a dans l'éducation. À la sortie du confinement, besoin d'une solution, BBB. Aujourd'hui, le MENJ figure au board (le conseil) et finance des features.

Bastien : Distinguer trois choses : le club utilisateur, le sponsoring de roadmap, le service.

Milovann : Ma problématique : je suis architecte de formation, je suis devenu libriste en 2004 et aujourd'hui, il y a plein d'open source dans le monde de l'architecture. J'ai lancé le projet [fordj.org](#) et je m'en sers pour travailler avec les étudiants, c'est une nouvelle façon de faire de l'architecture.

Vincent : Aujourd'hui, de part de ce que sait faire l'administration, c'est plus facile pour l'Etat d'avoir une relation commerciale (parce qu'on sait le faire) plutôt que d'avoir une relation avec des assos ou des fondations. Aujourd'hui, nous Grist, ça nous va bien que ce soit une boîte. Pour ce qui est des collectivités, elles seraient preneuse de Grist en SaaS.

Matthieu : Difficile pour une institution de trouver l'interlocuteur pertinent.

Bastien : Faire une base d'exemples harmonisé à consulter, avant même d'avoir une doctrine ?

Vincent : C'est compliqué de se détacher du besoin d'un contrat de support et de maintenance : sur Grist, on va avoir besoin de support, donc si des projets cherchent un modèle simple.

Pablo : Proposer une adhésion à une fondation est une chose facile.

* : Projets intéressants : <https://copiepublique.fr> et <https://www.clojuriststogether.org> qu'on pourrait vouloir faire passer à l'échelle. Voir aussi <https://github.com/bzg/opensource-challenges>.

Mercredi 14h - 15h code.gouv.fr Awesome : les projets à connaître

Le site internet est ici : <https://code.gouv.fr/fr/awesome/>

Pour pouvoir rentrer dans l'annuaire, il faut remplir les 4 critères : <https://github.com/codegouvfr/awesome-codegouvfr#awesome-codegouvfr>

Le score est déterminé depuis ces critères optionnels d'évaluation : <https://github.com/codegouvfr/awesome-codegouvfr#awesome-codegouvfr-score>

Les critères peuvent être remis en question.

Issues créées:

- <https://github.com/codegouvfr/awesome-codegouvfr/issues/21>
- <https://github.com/codegouvfr/awesome-codegouvfr/issues/22>
- <https://github.com/codegouvfr/awesome-codegouvfr/issues/23>
- <https://github.com/codegouvfr/awesome-codegouvfr/issues/24>
- <https://github.com/codegouvfr/awesome-codegouvfr/issues/25>

PR créées:

- <https://github.com/InseeFrLab/onyxia/pull/718>

Ressources:

- Outil pour générer un publiccode.yml rapidement: <https://github.com/codegouvfr/awesome-codegouvfr-tooling>
- S'appuyer sur le standard publiccode.yml: <https://github.com/publiccodeyaml/publiccode.yml>
- Formulaire en ligne pour obtenir un publiccode.yml: <https://publiccode-editor.etalab.studio/>

Mercredi 15h - 15h30 Atelier documentation pour code.gouv.fr

- Dépôt: <https://git.sr.ht/~codegouvfr/documentation>
- Site: <https://code.gouv.fr/documentation/>
- Référent: Bastien

Mercredi 15h - 16h Atelier Wikidata

Présentation de Wikidata, la base de connaissance des projets Wikimedia.

Lien vers un support de formation : [https://commons.wikimedia.org/wiki/File:D%C3%A9couvrir Wikidata - 2019.pdf](https://commons.wikimedia.org/wiki/File:D%C3%A9couvrir_Wikidata_-_2019.pdf)

Bilan et prospective

Pour clôturer ces deux jours, un temps de perspective sur l'avenir de la communauté BlueHats a débouché sur des propositions récurrentes de rencontres thématiques et/ou régionales, avec une volonté partagée de se revoir régulièrement pour pouvoir continuer à faire communauté.

En 2024, les participantes aimeraient que la communauté Bluehats :

- fasse d'autres événements comme celui-ci, hors Paris
- fasse 3 événements thématiques par 3 organisations différentes, axés sur la coopération et l'entraide
- fasse des hackathons ciblés avec des pitches qui puissent parler aux décideurs et aux financeurs
- soit un fond de financement de projets libres
- soit internationale
- soit un lieu incontournable et surtout inclusif pour incuber et échanger autour des logiciels libres

De la même manière, le bilan des deux jours était positif, les participants insistant sur les temps de rencontres et de collaboration, avec un équilibre entre le travail technique et des discussions collectives.

Les pépites deux participants à l'issue des deux jours :

- Content de pouvoir participer même sans être dev et de voir d'autres personnes hors de mon ministère. En revanche on est beaucoup d'hommes
- Je pars avec le fait de pouvoir bosser à nos trucs tout en aidant les autres
- Pas dev non plus, très content de voir que ça a pris et que le libre continue à se développer avec la Dinum
- Permis de mettre les pieds dans le domaine, hâte de vous écrire et de vous retrouver ailleurs
- Appris plein de choses sur la vie d'une communauté et d'un projet libre
- Rencontrer plein de gens que je ne connaissais pas +1
- Tous les sujets sur la soutenabilité. On a une promesse de soutenabilité et c'est se voir qui permet d'en discuter
- Dans la boule de neige, de voir la convergence des envies d'organiser d'autres événements
- Pour plus tard : j'attends avec impatience la liste de tous les contacts.