Document de travail

Projet de base de données encyclopédique mondiale

Global DB Encyclopedia

GDBE

Auteur : Erwan Courel

Suivi des modifications

|  |  |  |
| --- | --- | --- |
| Date | Personne | Action |
| 06/06/2015 | Courel | Création du document |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# But du document

Ce document a pour vocation de servir de base de travail pour le projet GDBE.

Il sert à définir les concepts entourant le projet, ainsi qu’à accompagner les phases de conception et d’éclairer les choix opérés lors de la réalisation du produit fini.

GDBE a pour but de devenir ce qui est communément appelé une « usine à gaz », et doit absolument être suivi et documenté pour rester cohérent sur le long terme.

Chaque modification significative du document amènera à la sauvegarde d’une copie de la précédente version, dans le but de conserver l’historique du projet.

# But du projet – Qu’est-ce que GDBE ?

Qu’est ce qui se cache derrière la notion de GDBE ?

GDBE est un acronyme pour Global DataBase Encyclopedia.

Le nom est un peu pompeux mais est relativement explicite : il s’agit de monter une encyclopédie en ligne, sous forme d’une gigantesque base de données, pour construire un modèle de données relationnel de l’ensemble des connaissances humaines.

## Pourquoi ?

Le but de ce projet n’est pas à caractère lucratif. Son but est plus « humaniste », dans le sens où il découle d’un triste constat. En 2015, et depuis la chute du bloc communiste, voire avant, le monde néo-libéral créé par les puissances occidentales – Américains en tête de gondoles, et Européens bien contents de ramer derrière – étouffe les peuples du monde entier, nous entraînant dans une spirale consumériste qui n’aura de fin que lorsque les conséquences sociales et/ou environnementales nous rattraperont, entraînant le déclin de la majorité des nations. Et par conséquent une dégradation sensible des conditions de vie de l’ensemble des populations. Dégradation qui n’a par ailleurs besoin d’attendre que le monde néo-libéral s’écroule, dans la mesure où le nombre de personnes passant sous le seuil de pauvreté ne cesse d’augmenter dans les pays occidentalisés, malgré un accroissement des richesses supérieur à l’accroissement de la population mondiale. Absurde n’est-ce pas ?

Ce constat global, bien que triste et je l’accorde, pour l’instant très péremptoire, ne doit pas nous conduire au fatalisme. Malgré l’apparente impuissance dans laquelle nous, peuples d’Europe mais aussi du monde entier, semblons placés, changer les choses est, je pense, toujours possible.

Afin de changer les choses de manière efficace, il convient de se poser la question qui nous sauvera : pourquoi ? « Pourquoi nous sentons-nous impuissants ? », « pourquoi le monde est-il dirigé tel qu’il est dirigé ? », « pourquoi les peuples ne se soulèvent pas outre mesure ? » sont autant de questions extrêmement intéressantes qui, par leur réponse et par l’apport de solution, peuvent nous sortir de cette situation. Le problème, aussi complexe soit-il, peut suivre au final le schéma suivant : « Problème 🡪 analyse 🡪 solution ».

Voici mon sentiment, la raison qui a fait que l’idée de ce projet m’est venu (car c’est bien le sujet du chapitre !) : Le monde est gouverné tel qu’il l’est parce que sa façon de gouverner rencontre l’indifférence des masses, ou son désintérêt de la vie politique. Contrairement à ce que j’ai pu dire plus tôt, la majeure partie de la population se sent moins impuissante qu’elle ne s’en fout ! Et que peut faire un gouvernement quand la majeure partie du peuple se désintéresse d’économie et de politique ? Tout.

Continuons alors à creuser : pourquoi une grande partie du peuple se désintéresse autant de la politique ? Il est possible de répondre de plusieurs manières :

* C’est chiant comme la pluie ;
* Les populations n’ont pas l’impression que le peu de démocratie à laquelle ils ont accès influe sur la politique du pays : « déficit démocratique » ;
* Les populations ne se sentent pas représentées par les politiques ;
* Les exactions des gouvernements, les interactions entre les pays et les instances supranationales ne sont pas clairement expliquées ou montrées du doigt : bref parlons de manque d’information.
* Les informations délivrées aux citoyens sont biaisées. Elles reflètent l’avis des gouvernements ou des multinationales tenant les chaînes d’information. Elles sont aussi et surtout partie prise. Le point de vue des médias occidentaux et des médias russes sont souvent diamétralement opposés. Qui vous avait dit que la guerre froide était finie ? Je crois qu’ici nous pouvons simplement parler de propagande.
* Par manque de temps.

Nous en sommes donc ici :

Sens du diagramme et légende :

Causes des causes 🡪 Causes 🡪 conséquences

Je ne prétends pas que le schéma ci-dessus est très beau, ni même exhaustif, loin s’en faut. Son but est toutefois de montrer d’où part mon raisonnement. Pour un schéma plus complet, mes seules connaissances, ni les connaissances de n’importe qui d’ailleurs, ne pourraient suffire : il s’agirait de mener un effort collectif, effort qui pourrait être porté grâce à une application telle que GDBE ! Sur ce schéma toujours, en jaune sont coloriés les deux points pour lesquels je souhaite apporter une valeur ajoutée.

## Comment ?

Aujourd’hui, une grande partie du savoir humain est présent sur la toile. Wikipédia est un parfait exemple de ce que peuvent permettre d’accomplir les nouvelles technologies : des outils communautaires, gratuits, connectés, soit complets, soit « complétables ». Toutefois, aussi utile que soit Wikipédia, il présente ses limites. C’est une encyclopédie en ligne comparable à une véritable encyclopédie. Des pages, et juste des pages. Les informations sont toutes (ou presque) présentées sous forme de texte brut. La notion de données est très peu présente.

Depuis quelques temps j’ai vu apparaître le projet Wikidata (toujours du groupe Wikimédia), dont le concept se rapproche peu ou prou de ce que je souhaite réaliser.

### Procédé et exemple

Le concept derrière la réalisation de GDBE est le suivant : Récupérer les informations où elles se trouvent (Wikipédia et autres), les formater, automatiquement ou semi automatiquement, sous des formats de données cohérents, et relier les informations entre elles. Par exemple : L’article sur Paris dans Wikipédia définit que Paris est une capitale. Une capitale est une ville. Et une ville est « un milieu physique où se concentre une forte population humaine, et dont l'espace est aménagé pour faciliter et concentrer ses activités : habitat, commerce, industrie, éducation, politique, culture, etc. ». D’après ces définitions, et une fois GDBE en place, toutes les requêtes incluant Paris, les capitales, les villes où les milieux physiques à forte population humaine se baseront en partie sur les données collectées sur la ville de Paris. Le concept peut être poussé plus loin : Paris est également une zone géographique. Une recherche lancée sur les zones géographiques concernera donc également les villes et tout autre élément qui aura été considéré comme une zone géographique !

### Mais qui construit ces liens ?

Les liens logiques entre les entités sont créés de deux manières. La première est automatique. Par analyse sémantique, GDBE peut construire les liens entre Paris, les capitales, les villes et les milieux à forte population humaine. Ce que GDBE ne pourra pas faire par contre, c’est inclure de lui-même un lien logique de type « Paris = zone géographique ». Une intervention humaine est donc nécessaire pour lui expliquer ce lien, pour qu’il puisse l’utiliser a posteriori. GDBE nécessite donc, comme Wikipédia, une communauté de rédacteur pour le faire grandir.

### Gestion de la subjectivité des contributeurs

Imaginons que je décide d’utiliser GDBE pour définir qu’un ananas, c’est un fruit mais aussi un canard. C’est évidemment absurde, mais quel procédé imaginer dans GDBE pour lutter contre des affirmations dont la légitimité peut plus difficilement être réfutée ? Par exemple, bien qu’athéiste et étant persuadé que le Dieu chrétien n’existe pas, je n’en ai pas de preuve formelle. L’information « Dieu existe » peut donc être difficile à démentir de manière factuelle.

#### Vérité et réalité

Nous entrons ici plus dans la philosophie que dans l’informatique. Qu’est ce qui est vrai, qu’est ce qui est réel ?

La vérité se définit globalement comme « l’adéquation entre la réalité et l’être humain qui la pense. » Cette définition est lourde de sens : cela signifie que la vérité est subjective. Il n’y a donc pas une vérité mais des vérités. Au moyen-âge, en Europe occidentale, la croyance en Dieu faisait plus partie du fait établi que du dogme. Dieu était donc une vérité, dans le sens où la (quasi ?) totalité de la population y croyait. De la même manière, si aujourd’hui le monde entier pensait que la Terre avait la forme d’une banane, alors le fait que la Terre a la forme d’une banane est la vérité. Mais cette vérité n’est pas nécessairement conforme à la réalité, comme le montre cet exemple.

Partant de cette base, on comprend rapidement que la réalité n’est pas objectivement atteignable dans une multitude de domaines.

#### Se limiter au factuel

GDBE n’a pas la prétention de déterminer ce qui est réel et ce qui ne l’est pas. Ceci ne peut pas être l’œuvre d’ailleurs de l’espèce humaine.

Toutefois il est toujours possible de se limiter au factuel pour les sujets portant à débat : Tout fait présent dans GDBE devra être sourcé/justifié sous peine d’être considéré comme un fait « discutable » et d’être filtré lors des utilisations de GDBE. Par exemple pour les croyances religieuses, l’existence ou l’inexistence du/des dieu(x) ne peut être prouvée de façon formelle, et la thèse comme l’antithèse seront donc de manière systématique estampillées « discutable ».

Toutes les sources/justifications seront également considérables comme « discutables » par la communauté.

#### Les faits empiriques

Tout ce qui a été présenté auparavant de fait pas mention des faits résultants de l’expérience, des faits qui sont considérés communément comme justes mais non prouvables.

Il peut s’agir par exemple d’axiomes mathématiques : 1+1 = 2. Ou par exemple de faits communément admis : le Canard a deux pattes. Rien ne nous dit que l’humanité ne se trompe pas depuis l’invention des mathématiques, et qu’en réalité 1 et 1 valent autre chose que 2…

#### La dénonciation d’un fait comme « discutable »

Supposons que vous tombiez dans GDBE sur une affirmation qui ne vous plaît pas, que vous jugez fausse par exemple. Une option pourra vous permettre de signaler que ce fait ne vous paraît pas conforme à la réalité. Si un nombre conséquent de personne fait la même chose (par exemple un millier. Ce nombre reste à définir plus tard) ce fait devient « discutable » et n’est plus considéré comme vrai.

S’ensuit deux cas :

* Si une large majorité de personnes (>95%, sur plus de 10 000 personnes par exemple) le déclarent comme discutable, alors le fait devient faux. Il est alors supprimé de GDBE, selon une procédure qui reste à définir.
* Si le fait entretien un débat (40% de vrai, 60% de faux par exemple), il pourra devenir intéressant d’explorer les causes des divergences d’opinion. Les divergences d’opinion deviennent factuelles, et sont donc incluses dans GDBE si le ratio ne devient pas extrêmement défavorable pour l’une des deux thèses (auquel cas nous basculerions dans le premier cas).

# Les grandes fonctionnalités

## L’analyse automatique de contenu

## Algorithmie pour l’enregistrement en base de données

## Interface de consultation des objets/concepts (diagrammes)

## Interface d’édition des objets/concepts – Faits, sourcing et avis

## Interface de consultation du modèle causal

## Interface de modification du modèle causal

## Interface de manipulation, d’extraction et de synthèse des données

## Interface de modèle prédictif

## Algorithmes de modifications de faits et de liens de causalités – Répercutions jusqu’en BDD

## How to/FAQ/Guide utilisateur