**PROGRAMME OPEN DU CNRS**

Fiche candidature

**A-Identité du logiciel**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Titre du projet** | Elvex | | | | | |
| **Acronyme** | Elvex | | | | | |
| **Porteur** | M. **☐** Mme **☐** | **Nom** : Clément | | **Prénom** : Lionel | | |
|  | **☐** Chercheur | | **☐** Ingénieur | | | |
| Enseignant chercheur | | **☐** Autre : précisez | | | |
|  | Courriel :  lionel.clement@u-bordeaux.fr | | Téléphone :  +33 (0)5 40 00 35 08 | |  | |
|  | | | | | | |
| **Institut principal du CNRS** | |  | **Délégation régionale** | | |  |
| **Unité (nom, numéro, adresse)** | |  | | | | |
| **Centre financier de l’unité (demander au gestionnaire du laboratoire)** | |  | | | | |
| **Contact SPV :** | |  | | | | |
| **Date de dépôt du dossier :** | |  | | | | |
| **Déclaration logicielle :** | | GNU GENERAL PUBLIC LICENSE Version 3, depuis 2019 | | | | |
| **Site web :** | | <https://github.com/lionelclement/Elvex> | | | | |

|  |
| --- |
| **Résumé du logiciel en 10 lignes (à destination des non-spécialistes du domaine):**  **Elvex est un logiciel de NLG (génération automatique de texte) qui n’utilise pas de donnée d’apprentissage, mais au contraire qui utilise des connaissances linguistiques pour produire du texte. Elvex traduit un *input* qui peut être un sens, mais aussi des intentions communicatives (comme un questionnement, un doute émis ou une injonction) en un ensemble de phrases qui respectent les contraintes syntaxiques de la langue.**  **Le premier intérêt de cette approche est de ne pas nécessiter de données massives ni de process d’apprentissage qui ont un coût important, autant économique qu’environnemental. Un autre intérêt, est de proposer des textes de façon déterministe et non à partir de données probabilistes. C’est-à-dire que l’on peut attester que les textes produits sont conformes aux *inputs* qui ont été donnés, pour autant que la grammaire et le lexique soient justes.**  **Le logiciel peut ainsi rapidement produire des textes correspondant à des emails, des lettres de publipostage, des descriptions de produits commercialisés, des exercices de langue, etc. avec l’assurance que le contenu est proposé dans une forme syntaxique correcte et sans contre-sens ou absurdité.** |
| **Objectif(s) du logiciel (3 lignes) :**  **Elvex permet de développer des applications de NLG. a) Qui ne nécessitent pas de donnée d’apprentissage manquantes ou couteuses, b) Qui assure la qualité de la production, c) Qui sont portées sur des matériels modestes** |
| **Valorisation envisagée après le programme OPEN (3 lignes) :**  **Il manque actuellement à Elvex un contexte de développement et de diffusion qui lui permettrait d’être valorisé auprès de la communauté des linguistes informaticiens et auprès des industriels.**  **Cela se résume à des éléments de visibilité comme une interface utilisateur et une documentation sérieuse, à la production de tests unitaires, et à l’écriture de données de tests.** |

**B-** **Description du logiciel**

**A – Contexte de développement du logiciel (<10 lignes)**

*Quels sont les raisons qui ont poussé à développer ce logiciel ? Dans quel cadre le logiciel a-t-il été développé (thèse, collaborations etc) ? Quels sont les développeurs du logiciel ?*

Mon travail de recherche en linguistique informatique m’a conduit à la réalisation d’un logiciel pédagogique d’analyse syntaxique (www.xlfg.org). Une rapide étude de la littérature m’a montré un manque en matière de NLG non stochastique qui pouvait être comblé par des méthodes et des formalisations linguistiques étudiées en développant XLFG. J’ai donc créé un formalisme et un algorithme pouvant répondre à certaines difficultés connues dans les approches de NLG (par exemple composer les entrées lexicales complexes comme les expressions idiomatiques, ou des constructions à verbe support qui sont difficilement produites ou impossibles à produire par les systèmes existants). L’expérimentation a été poursuivie par la réalisation du logiciel et son application à quelques cas d’école réputés difficiles.

**B- Description du logiciel actuel** **(<10 lignes)**

*Une description succincte du logiciel : comment le logiciel est-il utilisé ? Quels sont ses fonctionnalités principales ? Quels sont les avantages du logiciel ?*

Le logiciel actuel est un développement en C++ (34k lignes) qui permet de faire la preuve de concept. Il a été développé en respectant les règles de la programmation et a été expérimenté dans plusieurs situations réelles montrant sa bonne conception. Le logiciel est actuellement dans une version stable, mais il reste un projet de laboratoire difficilement accessible pour qui veut écrire une grammaire et un lexique sans avoir eu une communication rapprochée avec son concepteur.

**C-** **Utilisateurs du logiciel**

**C- Impacts actuels du logiciel (<10 lignes)**

*Qui utilise actuellement le logiciel et dans quel but ? Existe-t-il une communauté d’utilisateurs ? Sous quel licence le logiciel a été diffusé ? Comment a été motivé ce choix de licence ? Est-ce qu’il y a des développeurs du logiciel extérieurs au laboratoire ?*

Il n’existe actuellement aucune communauté d’utilisateurs grand public, mais quelques chercheurs et étudiants qui ont pu faire des expériences avec mon aide. Les buts sont pratiques (produire des textes automatiquement et de façon fiable) ou pour valider des expérimentations de recherche fondamentale (produire des corpus qui démontrent les hypothèses avancées et formalisées en Elvex).  
La licence est Open Source (GPL). Elle est motivée par le souhait de diffuser auprès de la communauté une production émise par l’Université. Il n’y a aucun développeur autre que son concepteur, c’est-à-dire moi-même.

**D- Perspectives du logiciel (<10 lignes)**

*A l’issue du programme OPEN, quel sera l’impact du logiciel et dans quel cadre ? L’utilisation du logiciel dans le cadre de collaborations ? La création d’un consortium d’utilisateurs ? La création d’une start-up ? etc.*

Tout est ouvert de ce point de vue, où il faut bien distinguer le logiciel en lui-même des données Elvex. En effet, le logiciel Elvex permet de faire tourner des grammaires Elvex pour réaliser des systèmes de génération de texte. Si le logiciel est en Licence Open Source, aucune contrainte de licence n’est porté sur les input et output d’Elvex. Nous pouvons faire une comparaison avec les compilateurs : certains sont en Open Source (gcc, g++, ocaml, etc.) et permettent de compiler des programmes d’autres licences. Ainsi l’impact du logiciel à l’issue du programme OPEN peut être le développement d’applications commerciales ou académiques utilisant diverses licences, libres ou non, mais utilisant le moteur Elvex pour faire du NLG dans le contexte industriel ou académique identifié.

**D-** **Développements futurs du logiciel**

**C- Besoins techniques (<10 lignes)**

*Quels besoins en développement sont identifiés ? Nouvelles fonctionnalités ? Industrialisation du code ? Est-ce qu’il y a des besoins matériels ?*

Il n’y a aucun besoin matériel identifié, mais un gros besoin en terme de travaux a) sur le code b) sur la documentation c) sur les données linguistiques. Ces travaux peuvent donner lieu à de nouvelles fonctionnalités identifiées par des linguistes ou des ergonomes relativement à la rédaction des grammaires et des lexiques et de l’interface utilisateur souhaitée. Cela peut aussi donner lieu à un gros chantier de codage pour optimiser et valider le logiciel sur des tests unitaires.

**C- Besoins autres (<10 lignes)**

*Aide à la diffusion du logiciel ? Accompagnement pour la réponse aux sollicitations extérieures (demandes industrielles) ? Besoins de structurations de la communauté (création de consortium etc.) ?*

Une grande partie de ce travail doit être valorisée par une communication scientifique. Cela suppose que les chantiers linguistiques (lexicographie, syntaxe, morphologie, etc) portant sur la formalisation Elvex doivent être publiés dans des revues et des conférences reconnues. La diffusion du logiciel lui-même suppose un travail de présentation autre qu’une simple entrée Github. La publication d’un site Web ou d’une application en lien avec les réseaux sociaux est indispensable à la valorisation du programme OPEN appliqué à Elvex.