



Jérémie ROQUET

127 Avenue du Général Leclerc, 54220 MALZÉVILLE

+33 (0) 6 33 07 69 95

<https://arkanosis.net/>
jroquet@arkanosis.net

Expert en ingénierie informatique

Ingénieur logiciel, recherche et développement, spécialisé dans le traitement de très gros volumes de données et l'optimisation logicielle, avec une expertise métier en traitement du langage naturel.

■ Compétences

Langages de programmation

- Expertise en C++, Python et ZSH
- Expérience avec de nombreux autres langages de programmation (dont D, Ruby, JavaScript, Lua, SED, AWK, Perl, C#, Java, GAS, NASM, LISP, PHP)
- Capacité à et volonté pour apprendre rapidement de nouveaux langages et de nouveaux paradigmes

Autres langages

- Expérience avec XML, YAML, JSON, (X)HTML (et accessibilité), CSS, SQL, plSQL, LaTeX...

Bibliothèques logicielles et frameworks

- Expertise avec la STL, Boost et la bibliothèque Python standard
- Expérience avec de nombreuses autres bibliothèques (dont jQuery, Tango, .NET, J2EE, J2SE, J2ME, requests, lxml, Django, Flask, SQLAlchemy, Bootstrap, Skeleton, GWT)

Bases de données

- Expérience avec MySQL, PostgreSQL and SQLite

Outils divers

- Expérience et / ou expertise avec de nombreux outils, dont :
 - ZSH
 - Emacs, Vim
 - Git, Subversion, Mercurial
 - Trac, Redmine, Bugzilla, Gerrit, Gitorious, Gitlab, Phabricator
 - Clang, GCC, MSVC
 - GDB, GProf, Perf, Valgrind, CodeAnalyst
 - SaltStack
 - TMux, Screen
 - Make, Scons...
- Recherche permanente de nouveaux outils pouvant simplifier le travail ou améliorer sa qualité

Systèmes d'exploitation

- Expertise en GNU/Linux
- Expérience avec Windows, {Open,Free,Net}BSD, Solaris, AIX, HP-UX, OSF/1

Expertise en

- Compression de données, optimisation logicielle, structures de données, gros volumes de données et passage à l'échelle, traitement du langage naturel...

Expérience avec

- Bioinformatique, compilateurs, intelligence artificielle, web, moteurs de jeux vidéos...

■ Expérience professionnelle

2009 – aujourd'hui

Exalead (Dassault Systèmes) — Ingénieur R&D logiciel

Exalead est un éditeur logiciel, leader dans la recherche en entreprise et les plates-formes pour le développement d'applications orientées recherche (SBA), filiale de Dassault Systèmes

- Développement d'un pipeline de traitement du langage naturel et d'analyse sémantique (détection de langue, reconnaissance d'entités nommées, annotation morpho-syntaxique, extraction de sentiments, correction orthographique, tokenisation des langues à composition, tokenisation des langues asiatiques...) avec une attention portée à l'évolutivité, aux performances et au passage à l'échelle. *Challenge : extraction d'un maximum d'informations utiles (bon rappel, bonne précision) en un minimum de temps à partir d'un maximum de données brutes non structurées*
- Développement et optimisation d'un nouvel index pour la recherche d'information (listes inversées, dictionnaires, compression, page cache...) avec une attention particulière portée à la fiabilité, aux performances et au passage à l'échelle — cet index a remporté l'édition 2012 des *3DS Innovation Forwards* de Dassault Systèmes. *Challenge : accès exhaustif et exact aux données recherchées en un minimum de temps et avec un maximum de flexibilité*
- Développement et optimisation d'un nouveau pipeline d'indexation pour la construction d'index en quasi-temps réel, avec une attention portée aux performances et au passage à l'échelle. *Challenge : indexation d'un maximum d'information en un minimum de temps (global), avec un minimum de latence (locale)*
- Design et développement d'une API mmap-like en userland pour Linux 2.6.32+ (potentiellement sans le syscall *userfaultfd*) pour l'utilisation transparente de systèmes de fichiers distribués (principalement HDFS) avec une attention particulière portée au nombre d'IOPS et à la latence. *Challenge : rendre l'utilisation de données stockées sur des systèmes de fichiers haute disponibilité aussi transparente que possible, en particulier du point de vue de la latence, sans nécessiter l'utilisation de modules kernel spécifiques*
- Outils et scripts pour les développeurs (dashboards, diagrammes, tâches automatisées, notifications, fonction shell d'auto-complétion...). *Challenge : simplifier le travail des développeurs, augmenter leur productivité et diminuer les erreurs humaines*

2007 – 2009

EPITA — Professeur assistant

EPITA est une école d'ingénieur en informatique et en technologies avancées

- Enseignement du développement logiciel et des langages de programmation aux étudiants de troisième année (C, C++, D, Ruby, SQL...), organisation et surveillance des examens...

Juillet 2008

INTI — Formateur technique

L'INTI est un institut de formation technique

- Enseignement de l'architecture logicielle, du langage C++ et de la programmation orientée objet à des développeurs logiciel

2007 – 2008

Eugen Systems (Ubisoft) — Développeur IA et game design

Eugen Systems est un studio de développement spécialisé dans les jeux de stratégie temps réel

- Développement du jeu de stratégie temps réel R.U.S.E. (game design, intelligence artificielle, moteur de son...), principalement en C++ et Python
- Outils pour les développeurs, les game designers et les artistes

Expérience extra-professionnelle

- 2007– aujourd'hui** **Wikipédia** — *Rédacteur, développeur d'outils et de bots, administrateur, masqueur*
Wikipedia est une encyclopédie libre en ligne
- Création et amélioration d'articles sur des sujets variés, protection contre le vandalisme
 - Développement et maintenance d'outils (JavaScript, Python, Lua...) pour les rédacteurs, la lutte contre le vandalisme et la détection de violations du droit d'auteur
 - Tâches de maintenance (entre autres de JavaScript, CSS)
- 2012– aujourd'hui** **Wikidata** — *Contributeur, développeur d'outils, administrateur*
Wikidata est une base de données libre, utilisée et éditée par l'homme et la machine
- Développement et maintenance d'outils (JavaScript) pour les contributeurs
 - Tâches de maintenance (entre autres de JavaScript)
- 2009 – aujourd'hui** **Shedskin** — *Contributeur*
Shedskin est un compilateur expérimental de Python vers C++
- Petites améliorations du compilateur
 - Implémentation et support de plusieurs bibliothèques standard de Python
- 2014 – aujourd'hui** **Regovar** — *Architecte et maître d'œuvre*
Regovar est un logiciel intégré d'analyse de données de séquençage haut débit
- Architecture initiale du projet
 - Recrutement du développeur principal
 - Maîtrise d'œuvre

Formation

- 2008 – aujourd'hui** **GAIC (NUCLAI) '08, FOSDEM '12, ESHG '15, C++FRUG '14 '15 '16, PyConFr '16**
Conférences et congrès (intelligence artificielle, logiciels libres, génétique, C++, Python)
- 2015** **Baylor College of Medicine de Houston & Université de Cologne**
Identification bioinformatique de gènes de maladies mendéliennes à partir de données NGS
- 2004 – 2009** **EPITA (École Pour l'Informatique et les Techniques Avancées)**
Titre d'expert en ingénierie informatique, spécialisation sciences cognitives et informatique avancée
- 2008 – 2009** **Université PARIS X NANTERRE**
Master II recherche en traitement automatique des langues
- 2004 – 2007** **Université de VERSAILLES SAINT-QUENTIN**
Licence en informatique (diplôme en cumulatif)

Langues

- Français** Langue maternelle
- Anglais** Courant (score de 960 au TOEIC en 2009)
- Allemand** Écrit et parlé, niveau scolaire
- Vietnamien** Notions avancées
- Chinois mandarin** Notions