

Deux services Gricad pour aider la recherche et l'enseignement

Transcription automatique et
documents de cours collaboratifs

Lucile Vadcard & Philippe Dessus
LaRAC, Univ. Grenoble Alpes
Atelier du 23 novembre 2023



1. TADDAM

- Pour "Transformations, analyses et développements de données et d'archives multimédias"
- Projet porté par la plateforme universitaire de données Grenoble Alpes et Patrick Juen (PACTE / GRICAD)
- Contact : Max Beligné
- Ce qui suit est inspiré de la séance du 26 juin 2023 par M. Béligné et P. Juen et de l'usage de l'outil

1. Outil de transcription de fichiers audio

[Conditions Générales d'utilisation et Mentions légale](#) TADAM

vadcardl 

Ajouter des fichiers audio à transcrire

Nommer votre tâche

Votre adresse mail

Déposer votre fichier ici (1Go max pour l'instant)

Format de sortie 

Récupérer une transcription si elle est terminée

Vadcard & Dessus | 2 services Gricad | Séminaire LaRAC, UGA | 23/11/23 | CC:BY-NC-SA

1. Fonctionnement de TADDAM

- URL : <https://intragere.univ-grenoble-alpes.fr/Tadddam>
- Se connecter avec identifiants UGA (y compris étudiants) – VPN si besoin
- Mentionner une adresse mail (UGA ou autre)
- Nommer la tâche
- Téléverser le dossier audio
 - Si > 5 h d'enregistrement, contacter les porteurs du projet
 - Drag & drop ➡ Bug
- Choisir le format de sortie
- « Lancer la transcription »
- **Ne rien faire** (ne pas cliquer sur "récupérer ma transcription") / on peut se déconnecter
- Premier mail de la part de sos-calcul-gricad de lancement de tâche : « Job is running »
- Deuxième mail de la part de sos-calcul-gricad de fin de tâche : « Job stopped normally »
 - Après quelques minutes ou quelques heures, selon la file d'attente des travaux ; 48h maxi
- Se reconnecter et « Récupérer la transcription »
- Reprendre le texte pour corrections

1. Récupérer la transcription

[Conditions Générales d'utilisation et Mentions légale](#)

vadcardl



Ajouter des fichiers audio à transcrire

Nommer votre tâche

Votre adresse mail

Déposer votre fichier ici (500Mo max pour l'instant)

Ajouter...

Format de sortie

SONAL: [Temps] <SPACE> [Locuteur] <SPACE> Texte



Lancer la transcription

Récupérer une transcription si elle est terminée

Degerli7

Récupérer la transcription



Degerli6

Récupérer la transcription



Degerli8

Récupérer la transcription



1. Format de sortie

- De + en + d'options
 - Différenciation des locuteurs
 - Indication temps
 - Formats compatibles
 - NVIVO
 - SONAL
 - ELAN
 - VTT pour sous-titres

des

Conditions Générales d'utilisation et Mentions Légales

vadcardl 

Ajouter des fichiers audio à transcrire

Nommer votre tâche

Votre adresse mail

Déposer votre fichier ici (1Go max pour l'instant)

Ajouter...

Format de sortie i

NVIVO: Temps <TAB> Locuteur <TAB> Texte

Texte brut (sans locuteur - ou fichier audio endommagé)

Locuteur <SPACE> Texte (sans horodatage)

Locuteur <TAB> Texte (sans horodatage)

Temps <SPACE> Locuteur <SPACE> Texte

NVIVO: Temps <TAB> Locuteur <TAB> Texte

ELAN: Temps <TAB> TempsFin <TAB><TAB> Locuteur <TAB> Texte

SONAL: [Temps] <SPACE> [Locuteur] <SPACE> Texte

Format VTT pour sous titrage (sans locuteur)

Récupérer une transcription si elle est terminée

Cliquez sur le bouton d'information pour en savoir beaucoup plus sur les fonctionnalités.







1. Qualité de la transcription

- Fonctionne préférentiellement en français
- Peut sauter des mots prononcés trop rapidement
- Moins bon sur vocabulaire spécifique
- Exemples d'erreurs
 - "Non, eux aussi" → Non, euh si
 - "quand on va être avec des pères en stage" → des pairs
 - "vous me parlez de l'adversité des intervenants" → la diversité
 - "je ne suis plus récultrice" → je suis puéricultrice
 - "si ce que vous manquez" → si ce qui vous manquait
 - "réanimation non-natale" → réanimation néo-natale

1. Remarques générales

- Outil interne à l'UGA – données protégées
- Rester raisonnable sur les volumes (ne pas enrayer le système)
- Supprimer les fichiers une fois récupérés
- Rester dans un usage d'appui à la recherche
- Principe général : au plus on sollicite au moins on est prioritaire
- Le système est en construction : évolutions et bugs
 - Relancer si besoin
- Principe fondamental : Travail  Consommation d'énergie...

2. La vie des enseignants : faire des ressources

- Enseigner implique de produire...
 - ... individuellement et collectivement...
 - ... un grand nombre de types **différents** de documents (cours, TP, TD, présentations, etc.),
 - ... destinés à des niveaux d'étudiants **différents**,
 - ... avec une structure **sophistiquée**, et des échéances de révisions rapprochées.

2. Économiser pour éviter de toujours réinventer la roue

- Beaucoup de temps est alloué à cette production, **pas toujours optimalement** : documents propriétaires avec de multiples versions, difficiles à modifier et convertir, forme souvent non harmonisée, etc.
- Utilité d'un système pouvant aider à produire, stocker, et diffuser ces documents

2. Des ressources libres et ouvertes ?

- de licence de partage **libre**
- de format de fichier **ouvert**
- **accessibles**/utilisables par tous (*responsive*)
- permettant une **évaluation** de ce qui est compris

2. Les langages de balisage léger

- L'objectif d'un langage de balisage léger est de faciliter la création de documents lisibles par les humains et également facilement analysables par les machines. Il est utilisé pour dans le cas où un langage de balisage complet comme le XML ou le HTML serait trop complexe.
- Utilisation du [Markdown](#) (étendu de quelques extensions) qui plus répandu et plus simple à appréhender pour les collègues.

2. Gitlab de Gricad

- L'outil Gitlab de Gricad, accessible par tout compte Agalan UGA, permet de
 - déposer des fichiers-sources en Markdown (présentations, documents de cours)
 - générer à la demande des fichiers HTML, PDF, ou epub, accessibles à tous
 - fichiers éditables par toutes les personnes autorisées

2. Caractéristiques principales des docs

- structurés sémantiquement, pouvant s'adapter à des styles différents
- modulaires & accessibles
- avec tous les **attributs** des cours (index, renvois croisés, images, vidéos, schémas, formules math, etc.)
- gérant des styles bibliographiques
- comportant des **QCM**
- pouvant admettre de nombreuses extensions (e.g., Graphviz)
- **éditables** collaborativement en asynchrone
- intégrables dans **toute plate-forme** d'enseignement (e.g., *Moodle*)

Portail de ressources à l'Inspé, UGA

Portail d'environ **200 documents de cours** (principalement sur le numérique et les sciences de l'éducation)

[Index des cours](#)

[Index des présentations](#)

- **Cours** plutôt théoriques avec QCM intégrés
- **Ateliers** (TD courts)
- **Tutoriels** (TP expliquant une procédure)
- **Ressources** (listes raisonnées et thématiques d'URL)
- **Syllabi** (programmes de cours)
- **Parcours** (pages pointant sur les ressources précédentes, par thème)

Démo

Merci pour votre attention !

Les personnes intéressées à contribuer à ce portail, ou à créer le leur, nous demandent !

Références

Pour plus d'informations :

- Dessus, P., & Besse, É. (2020). [Des ressources de cours libres et collaboratives pour une formation hybride des enseignants : Design et impact](#). *Distances et Médiations des Savoirs*, 31.
- Merci à [Naeka](#) pour avoir créé le système initial, et à Franck Pérignon, LJK, UGA, pour avoir permis le transfert dans le système actuel.
- Merci à Fabrice Ménard, DAPI, UGA, d'avoir porté le système sur Gricad et de le maintenir.