Vérificateur de liens (linkchecker)

Le service de vérificateur de lien scanne toutes les pages d’un site web pour identifier les liens hypertextes présents et vérifier leur validité. Il teste chaque lien pour s’assurer qu’il mène à une page web active et non à une page d’erreur, comme une erreur 404. Ensuite, il génère un rapport listant les liens cassés ou non fonctionnels.

- Linkchecker - Doc publique
  - Vérification des liens
  - Utiliser les résultats de vérification

Linkchecker - Doc publique

Vérification des liens

La présence de liens morts dans un site web entraîne une expérience utilisateur dégradée, et impacte négativement le classement des sites de l'UNIL dans les moteurs de recherche.

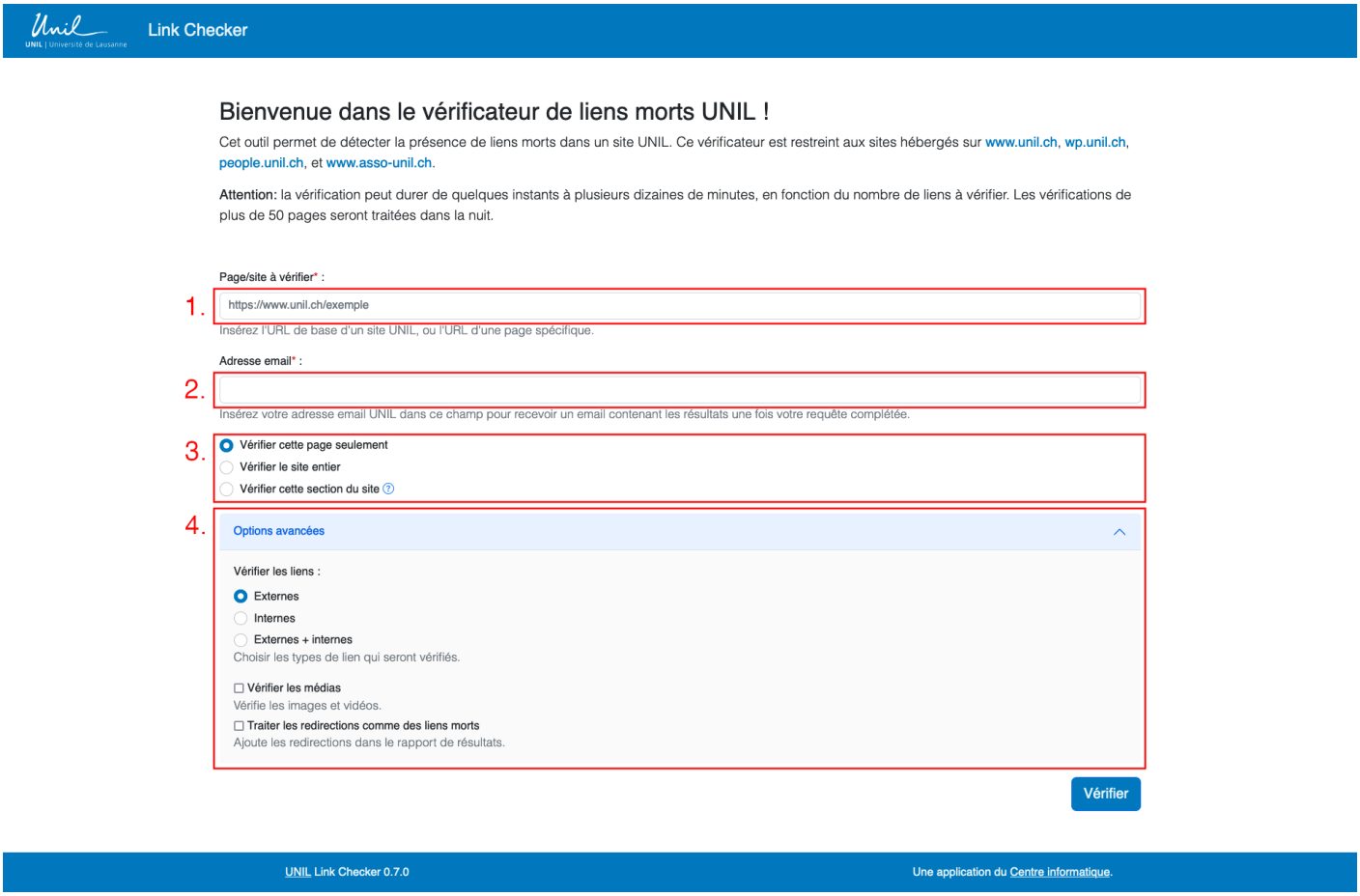
Trouver ces liens morts est cependant un travail laborieux et répétitif. C'est pourquoi nous proposons un outil permettant de vérifier automatiquement leur présence au sein des sites UNIL.

Cet outil est limité aux sites hébergés sur www.unil.ch, wp.unil.ch, www.asso-unil.ch, people.unil.ch . Il est également possible de vérifier un wiki en entier ou un livre (section) sur wiki.unil.ch (par contre, la vérification par section ne fonctionne pas par chapitre, en raison de la structure d'URLs du wiki)

Accéder au vérificateur de liens (seulement accessible depuis le réseau UNIL ou avec le VPN)

> Le vérificateur de liens ne peut vérifier que les pages publiques ouvertes sur Internet. Les pages privées ou protégées par mot de passe ne seront pas vérifiées.

Interface



1. Entrez l'**URL** du site (unil.ch/exemple) ou de la page que vous souhaitez vérifier. Si votre site utilise Wordpress, il est nécessaire d'indiquer le sous-domaine dans l'URL (wp.unil.ch/exemple, people.unil.ch/exemple, ou asso-unil.ch/exemple).
2. Entrez votre **adresse email UNIL**. Ce champ est obligatoire. Les résultats vous seront envoyés à cette adresse sous la forme d'une feuille Excel.
3. Définissez le **"périmètre" de recherche**: c'est-à-dire si l'outil va vérifier une page, une section, ou l'entièreté du site. La vérification par section cherchera dans toutes les pages du site dont l'URL commence par l'URL donné. (p.ex. si on donne l'URL "unil.ch/exemple/publications", la page "unil.ch/exemple/publications/sous-page" sera vérifiée, ainsi que toutes les autres pages dans la section "publications")
4. **Options avancées**:
    1. **Type de liens à vérifier**: cette option permet sélectionner quel types de liens seront vérifiés par l'outil. Par défaut, seuls les liens externes sont vérifiés. Il est également possible de vérifier les liens qui mènent à des pages internes au site vérifié (liens internes). Cette option est utile pour vérifier les éléments de menus, les références internes à d'autres pages, et les médias hébergés directement sur ce site. **Pour la plupart des cas, vérifier les liens internes est superflu, et ralentit significativement la vérification. Le cas le plus courant est la vérification des liens externes uniquement.**
    2. **Vérifier les médias**: permet de vérifier les images et vidéos qui pourraient se trouver sur le site. Si vous souhaitez vérifier des médias qui seraient hébergés directement sur le site à vérifier, il faut également sélectionner la vérification des liens internes, sinon seulement les images externes seront vérifiées. **Pour la**

plupart des cas, cette option est superflue, et ralentit significativement la vérification.

3. **Traiter les redirections comme des liens morts**: cette option fait remonter les redirections (statuts HTTP 3XX) dans la liste des résultats. Les redirections ont un impact négatif sur le référencement des sites web. Il convient de les éviter le plus possible et de toujours utiliser l'URL final si c'est faisable.
4. **Vérifier les pages orphelines (expérimental)**: cette option, qui ne fonctionne qu'avec la vérification complète ou d'une section d'un site, permet d'obtenir un rapport des pages orphelines de votre site web. Les pages orphelines sont des pages qui existent dans le site, mais qui n'est liée nulle part dans le site. ***Attention: cette option est expérimentale, et ne fonctionne pas pour l'heure avec les sites Jahia.***

> **Attention:** l'utilisation des options avancées ralentit la recherche ! Ne les utilisez que si vous voulez spécifiquement vérifier ces éléments.

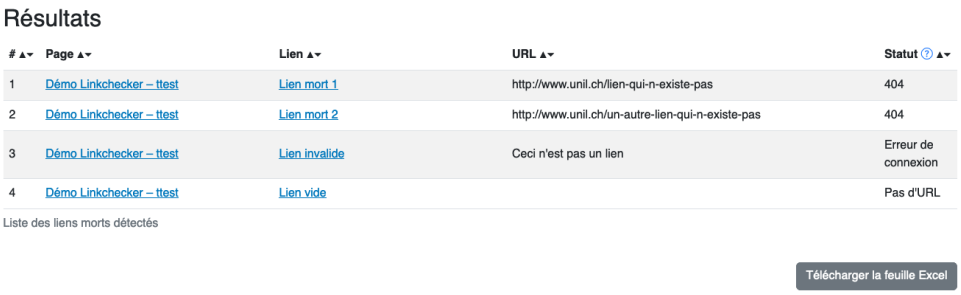
Lancez la vérification en cliquant sur le bouton "Vérifier".

La vérification se fait en trois phases:

1. Tous les pages à vérifier sont collectées. L'application affiche une roue animée.
2. Si le nombre de pages à vérifier est supérieur à 100, votre requête sera mise en queue et traitée pendant la nuit, afin d'éviter de surcharger les serveurs. Sinon, votre requête sera traitée à son tour (l'application ne traite qu'une seule requête à la fois)
3. Les liens présents dans chaque page sont collectés. Cette opération peut prendre de quelques instants à plusieurs minutes, en fonction du nombre de pages. L'application affiche une roue animée.
4. Les liens sont vérifiés un à un. Une barre de progression apparaît et l'avancement du processus est affiché. La durée dépend du nombre de liens à vérifier. Une estimation du temps restant est calculée et affichée, ainsi qu'une vue d'ensemble de la progression de la vérification.

> Si le nombre de pages à vérifier excède 100 pages, votre requête est gardée en mémoire et sera traitée dans la nuit. Vous recevrez les résultats par email. Sinon, la vérification se poursuit en temps réel.

Résultats



Les résultats de la vérification se présente sous forme de table, où chaque ligne représente un lien mort.

- **Page**: lien cliquable vers la page dans laquelle le lien mort se trouve.
- **Lien**: lien mort tel qu'il apparaît dans la page. Vous pouvez afficher la page morte en cliquant sur ce lien.
- **URL**: URL vers laquelle le lien mort renvoie.
- **Statut**: le statut d'erreur associé avec le lien mort. Ces codes peuvent vous aider à déterminer la raison pour laquelle il ne fonctionne plus.

Ces résultats vous sont également automatiquement envoyés par email à l'adresse entrée dans le formulaire

Domaines ignorés

Pour des raisons de performance, certains liens sont ignorés par défaut. C'est notamment le cas des liens internes, qui ne sont vérifiés qu'en activant l'option avancée, des services dont on sait qu'ils sont toujours joignables, et de ceux qui bloquent systématiquement le vérificateur.

Domaines internes à l'UNIL:

- agenda.unil.ch
- applications.unil.ch
- applicationsinter.unil.ch
- applicationspub.unil.ch
- av.unil.ch/hva
- editjahia.unil.ch
- fmpsrv01prd.unil.ch
- fmpsrv02prd.unil.ch
- fmpsrv03prd.unil.ch
- github.unil.ch
- ip.unil.ch
- jabba.unil.ch
- my.unil.ch
- news.unil.ch/display
- repopub.unil.ch
- sapcent.unil.ch

- serval.unil.ch
- unilogo.unil.ch

Domaines externes à l'UNIL:

- www.babla.fr
- doi.org
- nbn-resolving.org/urn:nbn:ch:serval
- researchgate.net
- linkedin.com
- nytimes.com
- www.cairn.info
- www.pexels.com
- pixabay.com
- instagram.com
- www.jstor.org
- onlinelibrary.wiley.com
- sagepub.com
- tandfonline.com
- sciencedirect.com
- oup.com
- dreamstime.com
- www.science.org/doi
- sciencemag.org
- www.embopress.org/doi
- journals.biologists.com
- web.archive.org
- maps.google.com

> Les publications scientifiques en ligne disposent souvent de protections additionnelles contre les robots. En conséquence, il arrive fréquemment que le vérificateur soit bloqué par ces sites, signalant le lien comme brisé. De tels faux-positifs sont malheureusement inévitables, et il convient de toujours vérifier les résultats. Pour cette raison, certaines publications scientifiques ne sont jamais vérifiées.

Utiliser les résultats de vérification

La présence de liens morts sur votre site peut nuire à l'expérience utilisateur et affecter votre référencement dans les moteurs de recherche. Voici quelques étapes simples pour vous aider à corriger ces liens et améliorer la qualité de vos pages.

Identifier les liens morts

Après avoir utilisé l'outil de vérification des liens, vous recevrez un rapport des liens morts sur vos pages. Chaque ligne du rapport contient des informations utiles :

- **Nom de la page concernée** : Le lien cliquable vers la page où se trouve le lien mort.
- **URL de la page concernée** : L'URL cliquable de la page où se trouve le lien mort.
- **Texte du lien brisé** : Le lien tel qu'il apparaît dans la page.
- **URL du lien brisé** : L'URL vers laquelle le lien renvoie.
- **Statut** : Le statut d'erreur associé au lien mort

Analyser les liens morts

Une fois que vous avez le rapport, examinez chaque lien mort :

- **404 - Page non trouvée** : Cela signifie que la page à laquelle le lien renvoie n'existe plus. Vérifiez si l'URL est correcte ou si la page a été déplacée ou supprimée.
- **401 ou 403 - Accès refusé** : Ces erreurs indiquent que vous devez être connecté pour accéder à la page ou que le vérificateur a été bloqué par le site en question. Vérifiez ces liens à la main, car ils peuvent toujours fonctionner.
- **301 ou 302 - Redirection** : Ces statuts indiquent que le lien redirige vers une autre page. Bien qu'il ne s'agisse pas nécessairement d'une erreur, des redirections multiples peuvent nuire à l'expérience utilisateur et à votre référencement. Essayez de trouver l'URL finale pour éviter de chaîner plusieurs redirections. Attention : les redirections figureront dans les résultats uniquement si l'option avancée "Traiter les redirections comme des liens morts" est activée.

Statuts d'erreurs

Les codes d'erreur HTTP nous informent sur la raison pour laquelle un lien ne fonctionne plus. Le plus commun est le code 404, indiquant que la page demandée n'a pas été trouvée. Une liste exhaustive des codes d'erreurs HTTP peut être trouvée sur Wikipedia.

Les statuts d'erreur 401 et 403 peuvent indiquer qu'une page requiert d'être connecté et autorisé pour la consulter, ou que le vérificateur a été bloqué par le site. Ces statuts demandent d'être vérifiés à la main.

Certaines autres erreurs peuvent se produire, et demandent souvent d'être vérifiées à la main :

Pas d'URL	Le lien a été défini sans URL.
URL invalide	L'URL contient une erreur (p.ex. espaces, caractères illégaux)
Non HTTPS	L'URL utilise le protocole HTTP au lieu de HTTPS. Le lien n'est pas à proprement parler mort, mais **la présence de liens non HTTPS a un impact négatif important sur le référencement par les moteurs de recherche**. Remplacez "http://..." par "https://...".
Syntaxe invalide	L'URL contient une erreur de syntaxe. La plus commune est la duplication du protocole: p.ex "https://https://exemple.com", qu'il convient de remplacer par "https://exemple.com". Lorsqu'on entre un nouveau lien dans Jahia, "https://" est déjà prérempli. Il est alors relativement fréquent de coller une URL dans le champs sans remplacer son contenu prérempli, causant cette erreur.
Certificat SSL invalide	Le certificat SSL du serveur à l'URL donnée est invalide. Cela peut arriver quand on essaie de joindre une page qui ne supporte que HTTP avec un URL qui spécifie "https://". Il est alors possible de le remplacer par "http://". Cependant, veuillez noter qu'**il est fortement déconseillé d'avoir des liens non HTTPS dans votre site**. Dans d'autres cas, ce sont des problèmes de configuration du serveur qui peuvent avoir plus ou moins d'impact sur la navigation de l'utilisateur (la page peut toujours être accessible). Dans ce cas, contactez l'administrateur du serveur concerné.
Erreur de connexion	Le serveur ne répond pas. Cela peut arriver si le serveur est hors-ligne, ou répond trop lentement. Ce problème peut être temporaire.
Trop de redirections	Ceci peut se produire s'il y a plus de 30 redirections avant d'arriver à la page finale. Utilisez si possible directement l'URL final.

Délai d'expiration dépassé	Le serveur prend trop de temps à répondre. La page peut cependant toujours être accessible.
Erreur inconnue	Erreur plus complexe.

Corriger ou supprimer les liens affectés

Voici les actions que vous pouvez entreprendre pour corriger les liens :

- **Modifier le lien** : Si la page a changé d'adresse, mettez à jour l'URL. Pour retrouver une page déplacée, visitez la page d'accueil du site concerné et utilisez sa barre de recherche si elle existe, ou bien recherchez le titre du lien ou des informations contextuelles dans un moteur de recherche. Vous pouvez ajouter le mot-clé `site:nomdusite.com` à votre recherche pour restreindre les résultats à ce site en particulier.
- **Remplacer le lien** : Si la page est introuvable, remplacez-le par un lien vers une autre source pertinente. Vous pouvez aussi vérifier s'il existe une archive du lien original sur l'Internet Archive, et substituer le lien mort avec le lien de son archive.
- **Supprimer le lien** : Si le lien ne peut pas être corrigé ou remplacé, supprimez-le complètement de votre page pour éviter de frustrer les utilisateurs et maintenir un bon référencement dans les moteurs de recherche.

Vérification des liens répétés

Si un lien mort apparaît plusieurs fois dans le rapport, cela peut indiquer qu'il se trouve dans des zones de votre site qui se répètent sur toutes ou partie des pages, telles que :

- **Menu**: Les liens présents dans la barre latérale peuvent être cachés dans un menu déroulant
- **Barre latérale** : Les liens présents dans la barre latérale peuvent être cachés dans un menu déroulant
- **Pied de page** : Le pied de page contient également des liens qui sont répétés sur toutes les pages

Les liens figurants dans ces parties de votre site sont cruciaux pour la santé de votre site, méritant d'autant plus d'attention.

Tester les changements

Après avoir effectué les modifications, il est crucial de tester tous les liens sur vos pages. Vous pouvez utiliser à nouveau l'outil de vérification des liens pour vous assurer que toutes les corrections ont été appliquées correctement et que les liens fonctionnent désormais.

Pour maintenir un site de qualité, effectuez régulièrement des vérifications de liens. Cela vous permettra de détecter et de corriger rapidement tout nouveau lien mort qui pourrait apparaître.