

□ Temps de lecture : 3 min.

Nous présentons la nouvelle plateforme donbosco.info, un moteur de recherche salésien conçu pour faciliter la consultation des documents liés au charisme de Don Bosco. Créé pour soutenir le Bulletin Salésien OnLine, il dépasse les limites des systèmes d'archivage traditionnels, souvent incapables de détecter toutes les occurrences des mots. Cette solution intègre un hardware et un software spécialement créés dans ce but, offrant également une fonction de lecture. L'interface web, volontairement simple, permet de naviguer parmi des milliers de documents en différentes langues, avec la possibilité de filtrer les résultats par dossier, titre, auteur ou année. Grâce à la numérisation OCR des documents PDF, le système identifie le texte même lorsqu'il n'est pas parfait, et adopte des stratégies pour ignorer la ponctuation et les caractères spéciaux. Les contenus, riches en matériel historique et de formation, visent à diffuser le message salésien de manière capillaire. Grâce au téléchargement libre de documents, on encourage l'enrichissement continu de la plateforme et on améliore ainsi la recherche.

Dans le cadre des travaux de rédaction du Bulletin Salésien OnLine (BSOL), la création de divers outils de support, dont un moteur de recherche, s'est avérée nécessaire.

Ce moteur de recherche a été conçu en tenant compte des limites actuelles dans les différentes ressources salésiennes disponibles sur le réseau. De nombreux sites proposent des systèmes d'archivage avec des fonctionnalités de recherche, mais ils ne parviennent souvent pas à identifier toutes les occurrences des mots, en raison de limitations techniques ou de restrictions introduites pour éviter la surcharge des serveurs.

Pour surmonter ces difficultés, au lieu de construire un simple archivage de documents avec une fonction de recherche, nous avons réalisé un véritable moteur de recherche, doté également d'une fonction de lecture. Il s'agit d'une solution complète, basée sur un hardware et un software spécialement mis en place dans ce but.

Lors de la phase de conception, nous avons évalué deux options : un software à installer localement ou une application *server-side* accessible via le web. Étant donné que la mission du Bulletin Salésien OnLine est de diffuser le charisme salésien au plus grand nombre de personnes, il a été décidé d'opter pour la solution web, afin de permettre à quiconque de rechercher et de consulter des documents salésiens.

Le moteur de recherche est disponible à l'adresse www.donbosco.info. L'interface web est

volontairement essentielle, « spartiate », afin de garantir une plus grande vitesse de chargement. La page d'accueil répertorie les fichiers et les dossiers présents, dans le but de faciliter la consultation. Les documents ne sont pas seulement en italien, mais également disponibles dans d'autres langues, sélectionnables via l'icône correspondante en haut à gauche.

La plupart des fichiers téléchargés sont au format PDF provenant de numérisations OCR (reconnaissance optique des caractères). Étant donné que l'OCR n'est pas toujours parfait, il arrive que tous les mots recherchés ne soient pas détectés. Pour pallier cet inconvénient, différentes stratégies ont été mises en œuvre : ignorer la ponctuation et les caractères accentués ou spéciaux, et permettre la recherche même en présence de caractères manquants ou erronés. Des détails supplémentaires sont disponibles dans la section FAQ, accessible en bas de page.

Étant donné la présence de milliers de documents, la recherche peut apporter un nombre très élevé de résultats. C'est pourquoi il est possible de restreindre le champ de la recherche par dossiers, par titre, auteur ou année : les critères sont cumulatifs et aident à trouver plus rapidement ce dont on a besoin. Les résultats sont classés en fonction d'un score de pertinence, qui tient actuellement compte principalement de la densité des mots-clés à l'intérieur du texte et de leur proximité.

Idéalement, il serait préférable de disposer des documents en format vectoriel plutôt que numérisés, car la recherche serait toujours précise et les fichiers seraient plus légers, avec des avantages conséquents en termes de vitesse.

Si vous possédez des documents en format vectoriel ou de meilleure qualité que ceux déjà présents dans le moteur de recherche, vous pouvez les télécharger via le service de téléchargement disponible sur www.donbosco.space. Vous pouvez également ajouter d'autres documents non présents dans le moteur de recherche. Pour obtenir les identifiants d'accès (nom d'utilisateur et mot de passe), envoyez une demande par e-mail à bsol@sdb.org.