🕐 Reading time: 3 min.

We present the new donbosco.info platform, which is a Salesian search engine designed to make it easier to consult documents related to the charism of Don Bosco. Created to support the Salesian Bulletin Online, it overcomes the limitations of traditional archiving systems, which are often unable to intercept all occurrences of words. This solution integrates dedicated hardware and specially developed software, also offering a reading function. The deliberately simple web interface allows you to navigate through thousands of documents in different languages, with the ability to filter results by folder, title, author, or year. Thanks to the OCR scanning of PDF documents, the system identifies the text even when it is not perfect, and adopts strategies to ignore punctuation and special characters. The contents, rich in historical and educational material, aim to spread the Salesian message in a widespread manner. With free uploading for documents, continuous enrichment of the platform is encouraged, improving the search.

As part of the work for the drafting of the Salesian Bulletin Online, it was necessary to create various support tools, including a dedicated search engine.

This search engine was conceived taking into account the limitations currently present in the various Salesian resources available online. Many sites offer archiving systems with search functionality, but often fail to locate all occurrences of words, due to technical limitations or restrictions introduced to avoid overloading the servers.

To overcome these difficulties, instead of building a simple archive of documents with a search function, we have created a real search engine, also equipped with a reading function. This is a complete solution, based on dedicated hardware and specially developed software.

During the design phase, we evaluated two options: software to be installed locally or a server-side application accessible via the web. Since the mission of the Salesian Bulletin Online is to spread the Salesian charism to the greatest number of people, it was decided to opt for the web solution, so as to allow anyone to search for and consult Salesian documents.

The search engine is available at www.donbosco.info. The web interface is

deliberately essential and "spartan" to ensure faster loading speeds. The "home page" lists the files and folders present, in order to facilitate consultation. The documents are not only in Italian, but also available in other languages, selectable via the icon at the top left.

Most of the uploaded files are in PDF format derived from scans with OCR (optical character recognition). Since OCR is not always perfect, sometimes not all the words searched for are detected. To overcome this, several strategies have been implemented: ignoring punctuation and accented or special characters, and allowing searches even in the presence of missing or incorrect characters. Further details can be found in the FAQ section, accessible from the footer.

Given the presence of thousands of documents, the search can return a very high number of results. For this reason, it is possible to narrow the scope of the search by folder, title, author, or year: the criteria are cumulative and help to find what you need more quickly. The results are listed based on a relevance score, which currently mainly takes into account the density of keywords within the text and their proximity.

Ideally, it would be preferable to have the documents in vector format instead of scanned, as the search would always be accurate and the files would be lighter, with consequent advantages in terms of speed.

If you have documents in vector format or of better quality than those already present in the search engine, you can upload them via the upload service available on www.donbosco.space. You can also add other documents not present in the search engine. To obtain access credentials (username and password), send a request via e-mail to bsol@sdb.org.