

روش‌های مختلف نمایه‌سازی و جستجو در متن

منظور از **سند** در اینجا، هر نوع متن اعم از سخنرانی، مصاحبه، یادداشت، مقاله، گفتار کوتاه و... است.

۱. نمایه‌زنی دستی پاراگرافی

در این روش هر **پاراگراف** در یک عبارت توصیفی خلاصه‌نویسی می‌شود. به این خلاصه **موضوع** می‌گوییم.

ذیل هر موضوع، کلمات متعدد ترکیبی تعریف می‌شود. به این ترکیب‌ها **نمایه** می‌گوییم.

هر نمایه به چند واژه تبدیل می‌شود. به این واژگان، **کلید واژه** می‌گوییم.

در نهایت موضوع‌ها، نمایه‌ها و کلید واژه‌ها، در نرم‌افزار ثبت می‌شود.

۲. جستجوی ساده ماشینی

امکان جستجوی ساده نمایه‌هایی که به صورت طبیعی در ترکیب هر سند وجود دارند و نیز کلید واژه‌هایی که در جملات وجود دارد را با نرم‌افزار فراهم می‌کنیم.

۳. جستجوی پیشرفته

امکان جستجوی نمایه‌ها و کلید واژه‌ها را به صورت عادی، عطفی و ترکیبی به صورت (با و بدون فلان واژه) را در مجموع اسناد فراهم می‌آورد.

۴. دستی مدخلی

در این روش، به پاراگراف‌ها کاری نداریم.

کل سند مورد نظر ماست.

در انتهای هر سند، نمایه‌ها و کلید واژه‌هایی را ثبت می‌کنیم.

۵. دستی مدخلی سلکتی

در این روش، قسمت‌های مهم هر سند را انتخاب کرده، نمایه‌ها و کلید واژه‌هایی را در کنار همان قسمت انتخاب شده، ثبت می‌کنیم.

۶. بارور کردن کلیدواژه‌ها (ترکیب ماشین و انسان)

۱. در این روش، ابتدا تمام کلید واژه‌ها را با کمک نرم‌افزار استخراج می‌کنیم.

۲. واژه‌های اضافی (اعم از تکراری‌ها، افعال، حروف ربط، ضمیرها، واژه‌های بی‌کاربرد و...) را حذف می‌کنیم.

۳. به صورت دستی، مترادف‌های هر کلیدواژه را در نرم‌افزار به همان واژه مرتبط می‌کنیم.

۴. به صورت دستی، متضادهای هر کلیدواژه را در نرم‌افزار به همان واژه مرتبط می‌کنیم.

۵. اعراب‌گذاری کلمات مشابه؛ مثل (سیر و سیر) و مثل (نفس و نَفَس)

۶. به صورت دستی، از ترکیب هر کلیدواژه با کلمات دیگر، نمایه می‌سازیم و در نرم‌افزار با همان کلیدواژه مرتبطش می‌کنیم.

۷. از نرم‌افزار می‌خواهیم که اگر پاراگرافی فاقد کلیدواژه بود آن را به ما نشان دهد تا در صورت لزوم برای آن به صورت دستی کلید واژه تعریف کنیم.

۸. ما مجموعه‌ای از واژه‌های بارور شده خواهیم داشت.

۷. جستجوی الاستیک

واژه‌های بارور شده را برای موتور جستجوی نرم‌افزار تعریف می‌کنیم. در این حالت با انجام عمل جستجو تمام عبارات حاوی واژه‌های مترادف، متضاد و نمایه‌های مرتبط با آن نمایش داده می‌شود.

۸. نمایه‌سازی کاربرمحور

این روش به دو شکل زیر انجام می‌شود:

۱. ارائهٔ سندها در فضای وب یا تحت شبکه به کاربران متعدد و ذخیرهٔ نتایج جستجوهای آنها

۲. درخواست از کاربران برای ثبت نمایه‌های پیشنهادی خود ذیل هر سند یا هر پاراگراف

۹. غنی‌سازی متون

در این روش ما تمام یا بخشی از اسناد خود را در اختیار تیم «قم‌نت» قرار می‌دهیم. آنها با علامت‌گذاری اصطلاحات موجود در مجموعهٔ اسناد، از طریق نرم‌افزاری که حاوی اصطلاحات شش «ویکی» در حوزهٔ علوم اسلامی است، مجموعه‌ای از اصطلاحات، نمایه‌ها و کلید واژه‌های موجود در سندهای ما را در اختیارمان می‌گذارند.

۱۰. هوش مصنوعی

هوش مصنوعی ابزار ماشینی و هوشمند برای تهیهٔ نمایه‌ها و کلیدواژه‌های موجود در سندهاست. عملکرد این ابزار محدود به واژگان مصرح در متن نیست بلکه به اجتهاد و تولید هوشمند نمایه از هر سند می‌پردازد.

این ابزار در حال حاضر در حوزهٔ علوم اسلامی به بلوغ نرسیده است. در حال حاضر، برخی مراکز نوع ضعیفی از آن را در اختیار دارند. اما تا رسیدن به بلوغ و اجتهاد و استفادهٔ کامل از آن فاصله داریم. ضمن اینکه برنامه‌نویسی آن بسیار پرهزینه و البته پیچیده و زمان‌بر است.

رضا صنعتی. اسفند ۱۴۰۰