

مقایسه جامعیت نسبی پایگاه‌های فارسی زبان مگ ایران، پایگاه اطلاعات علمی جهاد دانشگاهی (اس.آی.دی) و ایران مدکس در بازیابی اطلاعات حوزه دندانپزشکی

منصوره دمرچی لو^۱

محسن حاجی زین العابدینی^۲

چکیده

حجم مقاله‌های منتشر شده در حوزه علوم پزشکی نیز مثل سایر حوزه‌ها افزایش قابل توجهی یافته است و ارزیابی کارایی پایگاه‌هایی که به نمایه‌سازی مجله‌های این حوزه می‌پردازند، ضروری است. در حال حاضر، جستجو و بازیابی مقاله‌های فارسی در حوزه علوم پزشکی از طریق پایگاه‌های «مگ ایران»، «ایران مدکس» و پایگاه «اطلاعات علمی جهاد دانشگاهی» امکان‌پذیر است. از آنجا که جامعیت و مانعیت به عنوان دو معیار مهم برای ارزیابی نظام‌های بازیابی اطلاعات به کار می‌روند، این پژوهش با استفاده از روش پیمایشی و مقایسه‌ای، جامعیت نسبی سه پایگاه اطلاعاتی فوق در حوزه دندانپزشکی را بررسی نمود. به همین منظور ۲۹ کلیدواژه از حوزه دندانپزشکی انتخاب شد و پس از جستجو و تجزیه و تحلیل داده‌ها، نتایج نشان داد «ایران مدکس» دارای بالاترین میزان جامعیت با ۰/۷۲ است. همچنین، مشاهده شد که جامعیت نسبی پایگاه «مگ ایران» نسبت به SID بیشتر است.

کلیدواژه‌ها: پایگاه‌های اطلاعاتی، جامعیت نسبی، ایران مدکس، مگ ایران، پایگاه اطلاعات علمی جهاد دانشگاهی (اس.آی.دی)، دندانپزشکی.

.....
۱. دانشجوی کارشناسی ارشد علوم کتابداری و اطلاع رسانی پزشکی، دانشگاه علوم پزشکی ایران

Mansoureh.damirchi@gmail.com

۲. عضو هیئت علمی مرکز اطلاعات و مدارک علمی وزارت جهاد کشاورزی zabedini@yahoo.com

مقدمه

رشد و گسترش اینترنت، روشهای دستیابی به اطلاعات را دگرگون کرده و برای ایجاد سهولت و تسريع در مکان‌یابی اطلاعات، امکانات بسیاری در وب ایجاد شده است. پایگاه‌های اطلاعاتی که به نمایه‌سازی مقاله‌های مجله‌ها می‌پردازند، یکی از این ابزارها می‌باشند.

با توجه به اینکه هزینه‌های جستجو زیاد بوده و زمان قابل توجهی برای جستجوی اطلاعات در پایگاه‌ها توسط کاربران صرف می‌شود، مهم است که بدانیم کدام پایگاه‌ها بیش از همه مؤثرند. در حال حاضر، برای ارزیابی ابزارهای جستجوی اینترنت، معیارهای جدیدی ایجاد شده و مورد استفاده قرار می‌گیرد. یکی از این معیارها، ربط^۱ می‌باشد که برای ارزیابی نظامهای بازیابی اطلاعات و کارایی جستجوهای انجام شده، به کار می‌رود و به وسیله دو ملاک جامعیت^۲ و مانعیت^۳ سنجیده می‌شود (شاگری، ۱۳۸۷).

مانعیت، صحت یک جستجو را می‌سنجد. جستجو در صورتی به مانعیت کامل دست می‌یابد که هر پیشینه بازیابی شده در رابطه با یک پرسش، لزوماً به آن پرسش مربوط باشد و جامعیت مقیاسی از اثربخشی در بازیابی کلیه اطلاعات جستجو شده در یک پایگاه اطلاعاتی است. هنگامی جامعیت کامل محقق می‌شود که هر پیشینه منفرد که باید در رابطه با یک پرسش خاص یافت شود، حتماً ردیابی گردد (احمدی، ۱۳۸۵). به عبارت دیگر، جامعیت توانایی نظام در بازیابی کلیه مدارک مرتبط و مانعیت توانایی نظام در عدم بازیابی مدارک غیرمرتبط می‌باشد.

محاسبه مانعیت از طریق تقسیم تعداد مدارک مرتبط بازیابی شده به کل مدارک بازیابی شده در نتایج جستجو میسر است، اما برای محاسبه جامعیت باید علاوه بر تعداد مدارک مرتبط بازیابی شده، تعداد کل مدارک مرتبط موجود در پایگاه هم در دسترس باشد که حاصل آن از تقسیم تعداد مدارک مرتبط بازیابی شده به کل مدارک مرتبط موجود در پایگاه، به دست می‌آید.

-
1. Relevance.
 2. Recall.
 3. Precision.

حجم مقاله‌های منتشر شده در حوزه علوم پزشکی نیز همچون حوزه‌های دیگر افزایش قابل توجهی یافته است و به نظر می‌رسد ارزیابی کارایی پایگاه‌هایی که به نمایه‌سازی مجله‌های این حوزه می‌پردازند، ضروری است. طی بررسی‌های اولیه مشخص شد جستجو و بازیابی مقاله‌های فارسی در حوزه علوم پزشکی از طریق پایگاه‌های مگ ایران^۱، ایران مدکس^۲ و پایگاه اطلاعات علمی جهاد دانشگاهی^۳ ممکن است و از آنجا که جامعیت و مانعیت به عنوان دو معیار مهم برای ارزیابی نظام‌های بازیابی اطلاعات به کار می‌روند، این پژوهش به بررسی جامعیت نسبی سه پایگاه اطلاعاتی فوق در حوزه دندانپزشکی می‌پردازد.

اهمیت پژوهش

برای دسترسی به نتایج دقیق‌تر، کاربران باید با مزایای هر یک از پایگاه‌های اطلاعاتی آشنا شوند تا پایگاه مناسب را بدون اتلاف وقت، برای جستجو انتخاب کنند. نتیجه این پژوهش به متخصصان و دانشجویان حوزه دندانپزشکی کمک خواهد کرد تا پایگاهی را که برای بازیابی مقاله‌های فارسی در حوزه تخصصی آنها جامعیت بالاتری دارد، شناسایی کنند.

پیشینه پژوهش

«چشمه سهرابی» (۱۳۷۸) طی پژوهشی تأثیر استفاده از اصطلاحنامه را بر میزان جامعیت، مانعیت و مدت زمان جستجو در بانک‌های اطلاعاتی کتابشناختی، بررسی کرده است. ۷ بانک اطلاعاتی در ۶ گروه علوم انسانی، علوم پزشکی، علوم پایه، کشاورزی، فنی و مهندسی و هنر برای انجام پژوهش در نظر گرفته شد و در هر بانک اطلاعاتی، ۱۰ کلیدواژه به طور تصادفی و با کمک متخصصان انتخاب گردید. هر کلیدواژه، یک‌بار از طریق

-
1. www.magiran.com
 2. www.iranmedex.com
 3. www.sid.ir

اصطلاحنامه و بار دیگر با استفاده از زبان طبیعی، مورد کاوش قرار گرفت. میزان جامعیت، مانعیت و مدت زمان جستجوی اطلاعات بازیابی شده از طریق اصطلاحنامه و زبان طبیعی اندازه‌گیری شد. یافته‌های پژوهش نشان می‌دهد: نخست، میزان جامعیت نتایج بازیابی در بانکهای اطلاعاتی به هنگام استفاده از اصطلاحنامه ۲۴/۳۹٪ و به هنگام استفاده از زبان طبیعی ۸۹/۵۶٪ می‌باشد. بنابراین، استفاده از اصطلاحنامه باعث کاهش جامعیت می‌شود. دوم، میزان مانعیت نتایج بازیابی در بانکهای اطلاعاتی به هنگام استفاده از اصطلاحنامه ۰۶/۷۵٪ و به هنگام استفاده از زبان طبیعی ۴۶/۵۸٪ است. با وجود این، استفاده از اصطلاحنامه موجب افزایش میزان مانعیت جستجوها می‌شود. سوم، میانگین مدت زمان جستجوی نتایج بازیابی شده در بانکهای اطلاعاتی به هنگام استفاده از اصطلاحنامه ۴۴/۴ دقیقه و به هنگام استفاده از زبان طبیعی ۳۸/۶ دقیقه است. لذا استفاده از اصطلاحنامه باعث کاهش مدت زمان جستجو می‌گردد.

«ارشاد سرابی» (۱۳۸۰) ضمن بررسی و مقایسه نسخه‌های رایگان مدلاین^۱ در وب، میزان بازیافت و ضریب دقت شش نگارش متفاوت از مدلاین را که به طور رایگان در دسترس می‌باشند، بررسی کرده است. ۳۶ جستجو به زبان طبیعی و زبان مقید در فیلد عنوان انجام داده و در نتایج، تفاوت معناداری در بازیافت و دقت این نگارشات در زبان طبیعی و مقید ملاحظه نشده است. بر اساس نتایج، نگارشات مورد بررسی هنگام جستجو به زبان طبیعی، تعداد مدارک بازیافتی و ضریب دقت بالاتری را دارا هستند.

«اکبری و شهریاری» (۱۳۸۴) تاثیر اختلاف املاي انگلیسی - آمریکایی و بریتانیایی بر میزان جامعیت بازیابی اطلاعات را بررسی کرده‌اند. در این مطالعه، برای آنکه به صورت کمی تأثیر اختلاف املاي انگلیسی بر معیار جامعیت ارزیابی شود، معیار جامعیت کاوش برای چند واژه که دارای املاي متفاوت در انگلیسی آمریکایی و بریتانیایی هستند، در دو پایگاه اطلاع‌رسانی انگلیسی زبان (پاب مد^۲ و کب^۳) و دو پایگاه اطلاع‌رسانی

-
1. Medline.
 2. Pubmed.
 3. CAB.

فارسی - انگلیسی داخل کشور («پژوهشگاه اطلاعات و مدارک علمی ایران» و «ایران مدکس») که به نظام ذخیره و بازیابی فارسی - انگلیسی مجهز هستند؛ محاسبه و مقایسه شده است. در پایگاه اطلاع‌رسانی تخصصی علوم پزشکی (پاب‌مد)، که املائی مرجع آن انگلیسی - آمریکایی است، نتیجه جستجو با املائی آمریکایی و بریتانیایی هر واژه همپوشانی کامل دارد. از طرفی، در پایگاه اطلاع‌رسانی تخصصی علوم زیستی و کشاورزی (کب) که املائی مرجع آن انگلیسی - بریتانیایی است، چون کاوش از طریق اصطلاحنامه انجام می‌شود، در صورت ورود واژه با املائی آمریکایی، ابتدا جستجوگر به املائی بریتانیایی ارجاع داده می‌شود، و در نهایت امکان انجام جستجو با املائی بریتانیایی میسر می‌گردد. اما پایگاه‌های «پژوهشگاه اطلاعات و مدارک علمی ایران» و «ایران مدکس» به علت نادیده گرفتن تفاوت املائی دو گونه انگلیسی آمریکایی و بریتانیایی، نمی‌توانند مانع از دست رفتن اطلاعات مرتبط و کاهش جامعیت شوند.

«شاکری» (۱۳۸۷) طی پژوهشی میزان جامعیت و مانعیت ابزارهای کاوش اینترنت با واسطه جستجوی فارسی را بررسی کرد. ۱۰ ابزار کاوش دارای شرایط تحقیق بود که بررسی گردید و نتایج نشان داد موتور جستجوی وبگاه^۱ از نظر میزان مانعیت، بالاترین میزان و موتور جستجوی دهیو^۲ در رتبه دوم و گوگل فارسی در رتبه سوم قرار دارند. در ضمن راهنمای موضوعی آفتاب با مانعیت ۹٪، کمترین میزان مانعیت را داراست. علاوه بر این، جامعیت موتورهای مورد نظر نیز بررسی شد و مطالعه نشان داد که وبگاه بالاترین جامعیت را دارد و دهیو و گوگل فارسی به ترتیب در رتبه‌های دوم و سوم قرار دارند. راهنمای موضوعی آفتاب نیز با ۱۱/۶٪ کمترین میزان جامعیت را دارا می‌باشد.

«بار - آیلن»^۳ (۱۹۹۸) با استفاده از کلیدواژه Erdos، دقت و بازیافت نسبی شش موتور جستجو را بررسی کرد. وی ۱۰-۲۰ نتیجه اول را بررسی نمود و به این نتیجه رسید که دقت موتورها بالا، بازیافتشان پایین و همپوشانی آنها کم است.

-
1. www.webgah.com
 2. www.dahio.com
 3. Bar_ilan.

«سار»^۱ (۲۰۰۴) دقت و بازیافت موتورهای جستجو را هنگام استفاده از «قابلیت تصحیح کننده متون»^۲ بررسی کرد. نتایج پژوهش وی نشان داد استفاده از قابلیت تصحیح کننده متون، دقت را ۴٪ و بازیافت را ۱۱/۵٪ افزایش می‌دهد.

«شافی و رادر»^۳ (۲۰۰۵) جامعیت و مانعیت پنج موتور جستجو (سایروس^۴، هات بات^۵، گوگل، آلتاویستا و بایو وب^۶) را در زمینه بیوتکنولوژی بررسی کردند. نتایج نشان داد سایروس جامع‌ترین موتور جستجو در بازیابی اطلاعات علمی است و گوگل و هات بات در رتبه‌های بعدی قرار دارند. همچنین، تحقیقات نشان می‌دهد موتورهای جستجو مورد بررسی (بجز بایو وب) با زبان پرسش ساختار یافته خوب عمل می‌کنند، در حالی که بایو وب در مورد پرسشهای ساختار نیافته، بهتر عمل می‌کند.

«او گردی»^۷ (۲۰۰۷) طی پژوهشی، نتایج حاصل از جستجوی ژنها و پروتئینها را در پایگاه‌های اطلاعاتی «مرکز ملی اطلاعات بیوتکنولوژی»^۸ تجزیه و تحلیل کرد. مرکز ملی اطلاعات بیوتکنولوژی بخشی از کتابخانه ملی پزشکی ایالات متحده آمریکاست که بیش از ۳۰ پایگاه اطلاعاتی و دیگر منابع را از طریق وبسایت خود در دسترس قرار می‌دهد. روشهای مختلفی برای جستجو وجود دارد و با توجه به تکنیک جستجوی انتخابی، نتایج متفاوتی بازیابی می‌شود. معمولاً دو روش Enterz^۹ و BLAST^{۱۰} بیشتر از سایر روشها توسط پژوهشگران در موقعیتهای مختلف استفاده می‌شوند. (پژوهشگری که در حال

-
1. Sarr.
 2. Spellchecker.
 3. Shafi & Rather.
 4. www.scirus.com
 5. Hotbot.
 6. Bioweb.
 7. O' Grady.
 8. National Center for Biotechnology Information (NCBI).

۹. برنامه جستجوی کلیدواژه‌ای

۱۰. Basic Local Alignment Search Tool: برنامه‌ای که از مجموعه‌ای از سؤالهای آماده که شامل نام اسیدهای نوکلئیک و پروتئین می‌باشد، استفاد می‌کند و در بازیابی، رکوردهایی را که شبیه به سؤال می‌باشد، ارائه می‌دهد.

جستجوی اطلاعات درباره یک ژن بخصوص و مشخص است، احتمالاً از روش Enterze استفاده می‌کند؛ در حالی که پژوهشگری که به دنبال اطلاعاتی از یک ژن ناشناخته است، از BLAST استفاده می‌کند. هر یک از این روشها، تنظیمها و محدودیتهای مختلفی را برای جستجو پیشنهاد می‌کنند که می‌تواند توسط استفاده‌کننده تغییر داده شود. جستجوی ۷۰ ژن در سه پایگاه^۱ NCBI، انجام شد و جامعیت و مانعیت آنها محاسبه گردید. نتایج نشان داد با استفاده از هر یک از روشهای فوق و یا اعمال تنظیمهای مختلف، نتایج متفاوتی به دست می‌آید. همچنین، برخی نتایج نشان داد تنظیمهای پیش فرض برای جستجو، برای تمام جستجوها مطلوب نیست. این موضوع، یکی از نکات مهمی است که متخصصان اطلاع‌رسانی هنگام جستجو باید به خاطر داشته باشند و آن را به پژوهشگران منتقل کنند.

جامعه پژوهش

بررسیها نشان داد در حال حاضر ۴ پایگاه وجود دارد که مقاله‌های مجله‌های علوم پزشکی از طریق آنها قابل جستجو است. ایران مدکس، مگ ایران، پایگاه اطلاعات علمی جهاد دانشگاهی و پارس مدلاین^۲.

از آنجا که پارس مدلاین به علت مشکلات فنی قابل جستجو نبود، از جامعه پژوهش حذف گردید و سه پایگاه باقی‌مانده بررسی شد.

ایران مدکس، پایگاهی است که به نمایه کردن مقاله‌های نشریه‌های فارسی زبان حوزه علوم پزشکی می‌پردازد. شروع کار آن در سال ۱۳۸۲ بوده است و حدود ۱۵۰ مجله را در این حوزه پوشش می‌دهد.

مرکز اطلاعات علمی جهاد دانشگاهی تحت نظارت شورای علمی، خدمات خود را از طریق پایگاه اینترنتی www.sid.ir به عنوان بانک اطلاعات علمی کشور ارائه می‌دهد. این سایت حدود ۲۹۶ نشریه در زمینه پزشکی، علوم انسانی، علوم پایه، فنی و مهندسی،

.....
1. Gene database, RefSeq RAN collection, RefSeq Protein collection.
2. www.parsmedline.com

کشاورزی، هنر و معماری را نمایه می‌کند که ۱۰۴ عنوان از آنها، در حوزه علوم پزشکی است.

فعالیت سایت مگ ایران از سال ۱۳۸۰ آغاز گردیده و ضمن مکاتبه‌ها و رایزنی‌های مکرر با مدیران نشریه‌ها، تا کنون موفق به پوشش و ارائه خدمات به بیش از ۱۳۰۰ نشریه در حال انتشار شده است. این خدمات شامل درج شناسنامه نشریه، طرح روی جلد و فهرست مطالب هر شماره و اطلاعات کتابشناختی مقاله‌های آنها با استفاده از مندرجات نشریه است. این خدمات فقط با ارسال یک نسخه از هر شماره نشریه، به نشریه‌ها ارائه می‌شود و کاربران می‌توانند با استفاده از فهرست الفبایی، فهرست موضوعی و جستجو، به نشریه‌های مورد نظر خود دست یابند. این سایت، پایگاهی مرجع است که با هدف اطلاع‌رسانی و دسترسی به همه مجله‌های کشور توسط بخش خصوصی و به صورت مستقل اداره می‌شود و به هیچ اداره، سازمان، نهاد و شرکت دولتی و غیردولتی وابسته نیست. ۲۲۰ عنوان از مجله‌های این سایت در حوزه بهداشت و درمان است.

روش پژوهش

برای انجام این پژوهش از روش پیمایشی و مقایسه‌ای استفاده شده است. شیوه کار به این صورت بوده است که ابتدا با استفاده از اصطلاحنامه پزشکی فارسی، کلیدواژه‌ها مشخص شد. بررسی کلیدواژه‌های حوزه دندانپزشکی در ساختار درختی اصطلاحنامه پزشکی فارسی نشان داد ۵۱ کلیدواژه مرجع (انتخاب شده) در این حوزه وجود دارد که ۱۸ مورد آنها جزء واژه‌های پذیرفته شده اصلی هستند. ابتدا تصمیم گرفته شد جستجو با استفاده از ۱۸ کلیدواژه مورد نظر انجام گیرد، اما در هنگام انجام پژوهش با این روش، مشکلاتی پیش آمد که مجبور به تغییر کلیدواژه‌ها شدیم. برای نمونه، در مورد کلیدواژه «ارتدَنسی» که یکی از اصطلاحات پذیرفته شده اصلی در این حوزه است، کلیدواژه‌های «دستگاه‌های متحرک ارتدَنسی» و «ارتدَنسی اصلاحی» که از اصطلاحات اخص «ارتدَنسی» هستند و یا کلیدواژه «دستگاه‌های ارتدَنسی» که اصطلاح مرتبط آن است و انتخاب شده نیز

می باشند، در نظر گرفته نشده اند و فقط کلیدواژه «ارتدَنسی» در نظر گرفته شده است. از آنجا که نمایه سازی هیچ یک از این سه پایگاه با استفاده از زبان کنترل شده و یا اصطلاحنامه پزشکی فارسی انجام نمی شود، مشاهده شد که در موارد بسیاری، نتایج جستجو با کلیدواژه منتخب نسبت به نتایج جستجو با کلیدواژه انتخاب نشده، کمتر است. از سویی، گاه مشاهده شد که کلیدواژه‌هایی در اصطلاحات رایج میان متخصصان حوزه دندانپزشکی وجود دارد، که طبعاً در عنوانهای مقاله‌های تألیف شده توسط آنان نیز به کار می‌رود، ولی در اصطلاحنامه (در کلیدواژه‌های انتخاب شده و انتخاب نشده) ذکر نشده‌اند. در نهایت، تصمیم گرفته شد کلیدواژه‌های انتخابی برای انجام پژوهش به قرار زیر باشند:

علاوه بر ۱۸ مورد فوق (کلیدواژه‌های اصطلاحنامه)، اصطلاحات انتخاب نشده آنها در اصطلاحنامه و اصطلاحاتی که در اصطلاحنامه نیامده است ولی به تجربه مشاهده شده که در میان اهل فن متداول است نیز به جامعه پژوهش افزوده شد. برای مثال، در مورد کلیدواژه «ارتدَنسی»، اصطلاح انتخاب نشده آن در اصطلاحنامه یعنی «ارتودانتیکس» و اصطلاح پیشنهادی آن (ارتودَنسی) نیز جستجو شد. در نهایت، مجموع کلیدواژه‌های مورد جستجو، ۲۹ کلیدواژه شد. (فهرست کلیدواژه‌های مورد استفاده در پیوست ذکر شده است).

همچنین، ذکر این نکته خالی از فایده نیست که اگرچه کلیدواژه دندانپزشکی در اصطلاحنامه با املای «دندان‌پزشکی» پیشنهاد شده است اما ملاحظه شد که در صورت استفاده از این املا، در پایگاه‌های SID و Magiran هیچ موردی بازیابی نمی‌شود و در Iranmedex نیز مقاله‌های بازیابی شده با املای «دندانپزشکی» نسبت به «دندان‌پزشکی» به طور قابل توجهی بیشتر بود. به همین دلیل در تمام موارد، با وجود پیشنهاد اصطلاحنامه، املاي «دندانپزشکی» استفاده شده است.

جستجوی ۲۹ کلیدواژه تعیین شده، به منظور اجتناب از تغییراتی که ممکن است به علت روزآمدسازی پایگاه‌ها پیش بیاید، در طی یک روز در سه پایگاه مگ ایران، ایران مدکس و پایگاه اطلاعات علمی جهاد دانشگاهی انجام شد. هر یک از جستجوها به طور

جداگانه بررسی و تعداد مدارک مرتبط در هر مورد یادداشت گردید. بررسی نتایج جستجوها نشان داد میان نتایج حاصل از پایگاه‌ها همپوشانی وجود دارد؛ بدین معنا که در اکثر جستجوها، مدارکی مشاهده شد که در میان پایگاه‌ها مشترک می‌باشند. لذا برای تعیین تعداد مدارک مشترکی که میان آنها وجود دارد، نتایج حاصل با هم مقایسه گردید؛ بدین ترتیب که مثلاً نتایج حاصل از جستجوی کلیدواژه «آسیب‌شناسی دهان» در پایگاه ایران مدکس، یکبار با نتایج حاصل از همین جستجو در پایگاه مگ ایران و یکبار با نتایج حاصل از همین کلیدواژه در پایگاه SID مقایسه شد و در پایان داده‌ها برای دستیابی به نتایج تجزیه و تحلیل گردید.

سنجش جامعیت

منظور از جامعیت یا بازیافت، توانایی یک سیستم در بازیابی تمام یا بیشتر مدارک مرتبط در پایگاه است. بنابراین، برای محاسبه مقدار آن نه تنها تعداد مدارک مرتبط بازیابی شده، بلکه تعداد مدارک مرتبط بازیابی نشده را نیز باید داشته باشیم. با این توصیف، به نظر می‌رسد محاسبه جامعیت یک پایگاه تقریباً غیرممکن باشد. چون نمی‌توان به تعداد مدارک مرتبطی که به دلایلی در بازیابی‌ها ملاحظه نمی‌شود، دست یافت. «کلارک و ویلت» (۱۹۹۷ ذکر شده در شافی، ۲۰۰۵) روشی برای محاسبه جامعیت نسبی^۱ در محیط وب به کار برده‌اند که در این مقاله نیز از روش آنها برای محاسبه جامعیت نسبی پایگاه‌های مورد نظر استفاده شده است. آنان، جامعیت نسبی را به این صورت تعریف می‌کنند:

جامعیت نسبی:

تعداد کل مدارک مرتبط بازیابی شده توسط یک پایگاه
تعداد کل مدارک مرتبط بازیابی شده توسط کلیه پایگاه‌های مورد مطالعه

1. Relative recall.

بدین ترتیب که اگر سه پایگاه مورد نظر را با حروف a, b و c نامگذاری کنیم، آنگاه مدارک بازیابی شده به وسیله آنها $a1$, $b1$ و $c1$ خواهد بود. اگر بین پایگاه‌ها همپوشانی وجود نداشته باشد، جامعیت نسبی حاصل تقسیم $a1$ بر $a1+b1+c1$ خواهد بود و اگر بین پایگاه‌ها همپوشانی وجود داشته باشد، و برای مثال همپوشانی پایگاه a و b، $b2$ نامگذاری شود؛ جامعیت نسبی پایگاه a برابر خواهد بود با: $a1$ در صورت و $a1+b2+c2$ در مخرج.

تجزیه و تحلیل داده‌ها

همان‌طور که در روش پژوهش اشاره شد، بررسی نتایج نشان داد میان مدارک بازیابی شده از پایگاه‌ها همپوشانی وجود دارد. جدول شماره ۱، تعداد مدارک بازیابی شده و جدول شماره ۲، مدارک مشترک میان پایگاه‌ها را به تفکیک نشان می‌دهند.

جدول ۱. تعداد مقاله‌های بازیابی شده توسط پایگاه‌ها

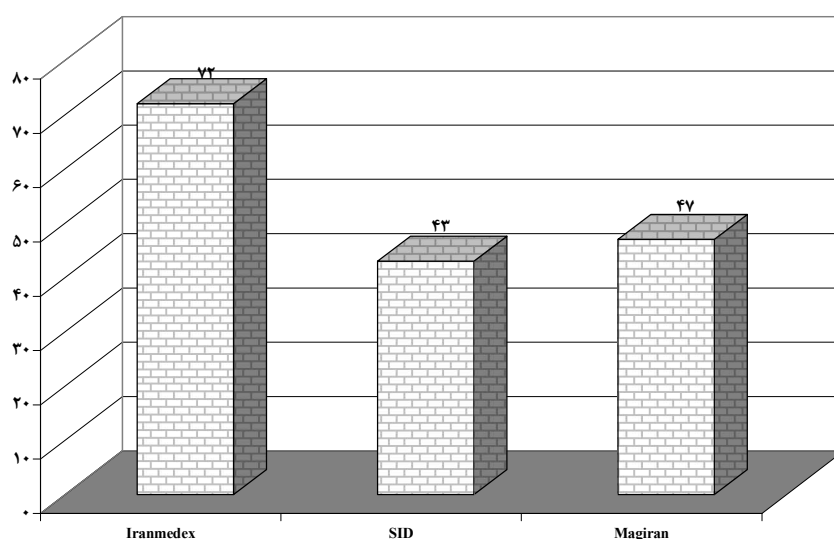
نام پایگاه	تعداد مقاله‌های بازیابی شده	تعداد مقاله‌های مرتبط بازیابی شده
Iranmedex	۷۸۳	۵۳۰
SID	۱۷۸	۱۷۳
Magiran	۱۷۷	۱۶۹

جدول ۲. تعداد مقاله‌های مشترک میان پایگاه‌ها

نام پایگاه‌ها	Iranmedex & Magiran	Iranmedex & SID	SID & Magiran
تعداد مقاله‌های مشترک	۸۶	۱۲۳	۱۰۶

با توجه به همپوشانی موجود، برای محاسبه مقدار جامعیت نسبی هر پایگاه، از فرمول دوم استفاده شد. به عبارت دیگر، پس از تعیین تعداد مدارک مرتبط بازیابی شده هر پایگاه، تعداد مدارک مشترک بازیابی شده در میان آنها تعیین و سپس محاسبه انجام گردید.

تجزیه و تحلیل داده‌ها نشان می‌دهد ایران مدکس دارای بالاترین میزان جامعیت، معادل ۰/۷۲ می‌باشد. جامعیت نسبی پایگاه مگ ایران ۰/۴۷ و پایگاه اطلاعات علمی جهاد دانشگاهی ۰/۴۳ به دست آمد. (نتایج در نمودار شماره ۱، قابل مشاهده است.)



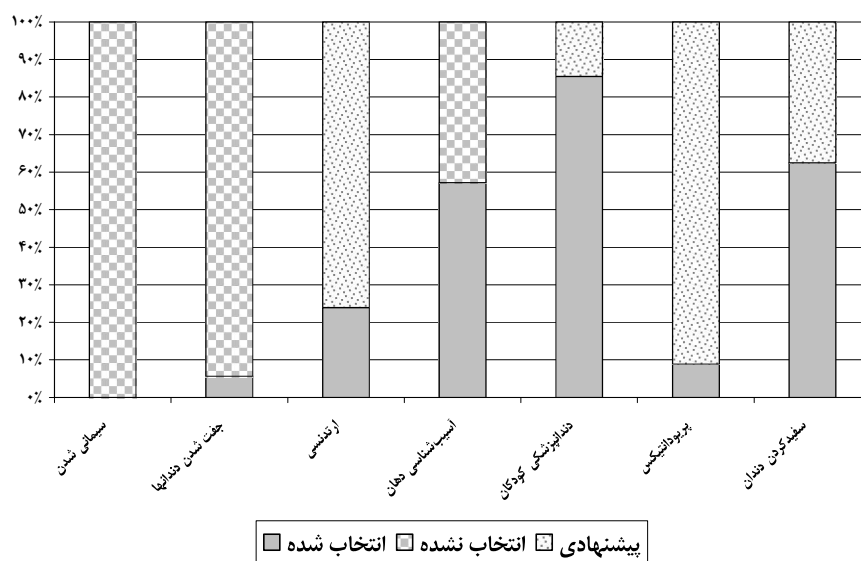
نمودار ۱. مقایسه جامعیت نسبی پایگاه‌ها

همان‌طور که در پیوست مشاهده می‌شود، در میان ۱۸ کلیدواژه اولیه‌ای که از اصطلاحنامه انتخاب شده بود، ۷ مورد دارای اصطلاحات نامرجح (ذکر شده در اصطلاحنامه یا پیشنهاد شده توسط پژوهشگر) می‌باشند که به منظور انجام مقایسه‌ای میان نتایج حاصل از جستجو توسط کلیدواژه‌های مرجح و نامرجح اصطلاحنامه، از این ۷ کلیدواژه استفاده شد (جدول ۳). کلیدواژه‌های مرجح به صورت **bold** و کلیدواژه‌های نامرجح در زیر آنها قید شده است. کلیدواژه‌های زیرخط دار مواردی هستند که توسط پژوهشگر پیشنهاد شده‌اند و در اصطلاحنامه به صورت انتخاب شده یا انتخاب نشده وجود ندارند.

جدول ۳. فهرست هفت کلیدواژه‌ای که دارای اصطلاحات نامرجح هستند

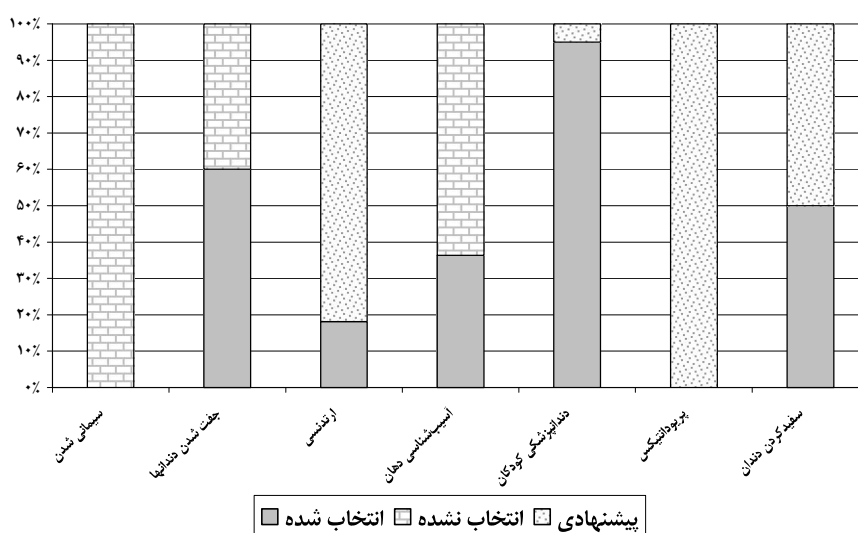
سیمانی کردن • سمان • سمنت	جفت شدن دندانها • اکلوزن دندان	ارتدنسی • ارتودانتیکس • <u>ارتودنسی</u>	آسیب شناسی دهان • پاتولوژی دهان
دندانپزشکی کودکان • پدودانتیکس • <u>دندانپزشکی اطفال</u>	پریودانتیکس • پریودنتیکس • <u>پریودنتولوژی</u>	سفید کردن دندان • <u>بلیچینگ</u>	

ابتدا مقایسه نتایج حاصل از بازیابی کلیدواژه‌های فوق در هر یک از پایگاه‌ها به طور جداگانه (نمودارهای ۲، ۳ و ۴) انجام شد. در محور افقی فقط به ذکر کلیدواژه مرجح اکتفا شده و میزان نتایج جستجوی اصطلاحات انتخاب شده، انتخاب نشده و پیشنهادی هر کلیدواژه به تفکیک نشان داده شده است.



نمودار ۲. مقایسه نتایج جستجو میان کلیدواژه‌های انتخاب شده، انتخاب نشده و پیشنهادی در پایگاه ایران مدکس

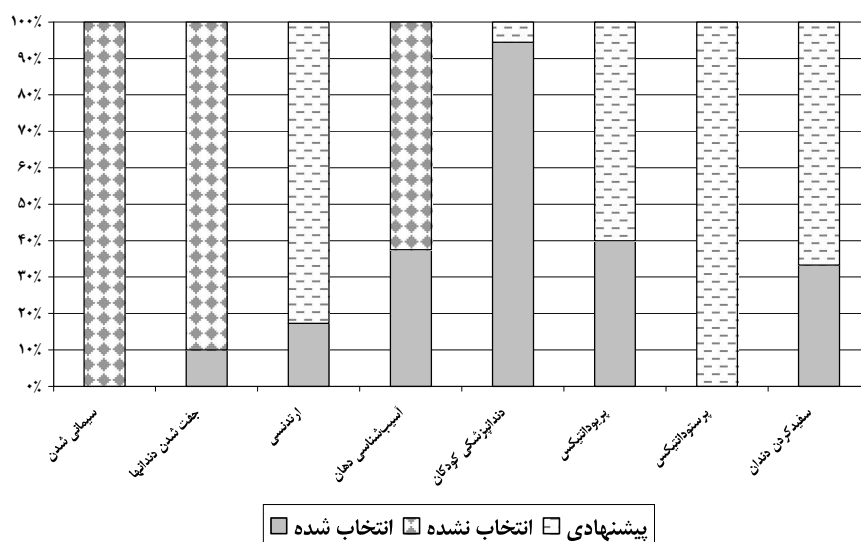
همان‌طور که در نمودار ۲ مشاهده می‌شود، ۱۰۰٪ نتایج حاصل از جستجوی کلیدواژه «سیمانی کردن» مربوط به اصطلاحات نامرجح است. در مورد «پریودانتیکس»، «ارتدنسی» و «جفت شدن دندانها» نیز درصد قابل توجهی از نتایج (به ترتیب ۹۳٪، ۷۶٪ و ۹۴٪) به اصطلاحات نامرجح مربوط می‌باشد. نتایج نشان می‌دهد که تنها در مورد کلیدواژه‌های «دندانپزشکی کودکان»، «سفید کردن دندان» و «آسیب‌شناسی دهان» درصد بیشتری از نتایج (به ترتیب ۸۵٪، ۶۲٪ و ۵۷٪) متعلق به اصطلاح انتخاب شده می‌باشد.



نمودار ۳. مقایسه نتایج جستجو میان کلیدواژه‌های انتخاب شده، انتخاب نشده و پیشنهادی در پایگاه مگ ایران

مقایسه نتایج در پایگاه مگ ایران نشان می‌دهد (نمودار ۳) ۱۰۰٪ نتایج حاصل از جستجوی کلیدواژه‌های «سیمانی کردن» و «پریودانتیکس» به اصطلاحات نامرجح مربوط می‌باشد. در مورد «ارتدنسی» و «آسیب‌شناسی دهان» نیز به ترتیب ۸۲٪ و ۶۳٪ نتایج مربوط به جستجوی کلیدواژه‌های نامرجح است. ۹۵٪ نتایج در جستجوی کلیدواژه «دندانپزشکی کودکان» و ۶۰٪ نتایج در مورد «جفت شدن دندانها» به اصطلاح مرجح مربوط می‌باشد. در

مورد کلیدواژه «سفید کردن دندان» ۵۰٪ نتایج مربوط به کلیدواژه مرجح و ۵۰٪ به کلیدواژه نامرجح مربوط است.



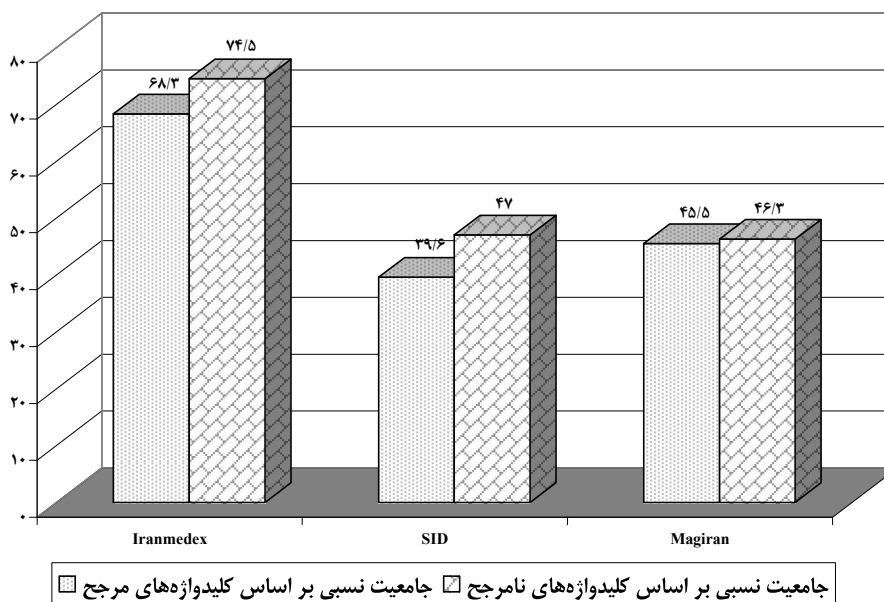
نمودار ۴. مقایسه نتایج جستجو میان کلیدواژه‌های انتخاب شده، انتخاب نشده و پیشنهادی در پایگاه اطلاعات علمی جهاد دانشگاهی

همان‌طور که نمودار ۴ نشان می‌دهد، ۱۰۰٪ نتایج جستجوی کلیدواژه «سیمانی کردن» در پایگاه SID مربوط به اصطلاحات نامرجح است. نتایج نشان می‌دهد تنها در مورد کلیدواژه «دندانپزشکی کودکان» اصطلاح مرجح (۹۴٪) نتایج بیشتری را به خود اختصاص می‌دهد و بخش اعظم نتایج جستجوی «جفت شدن دندانها» (۹۰٪)، «ارتدنیسی» (۸۳٪)، «آسیب شناسی دهان» (۶۲/۵٪)، «پریودانتیکس» (۶۰٪) و «سفید کردن دندان» (۶۷٪) به اصطلاحات نامرجح مربوط است.

مقایسه نتایج حاصل از سه نمودار فوق نشان می‌دهد نتایج حاصل از کلیدواژه «سیمانی کردن» در هر سه پایگاه مربوط به اصطلاحات نامرجح است و تنها در مورد

کلیدواژه «دندانپزشکی کودکان» بخش قابل توجهی از نتایج در هر سه پایگاه مربوط به اصطلاح مرجع می‌باشد.

به طور کلی می‌توان گفت، نتایج حاصل از جستجوی اصطلاحات نامرجح بیشتر از نتایج حاصل از جستجوی اصطلاحات مرجع می‌باشد. برای بررسی دقیق‌تر، جامعیت نسبی پایگاه‌ها یکبار بر اساس نتایج کلیدواژه‌های مرجع هفت مورد فوق و یکبار بر اساس نتایج کلیدواژه‌های نامرجح آنها (انتخاب نشده و پیشنهادی) محاسبه شد. بر اساس یافته‌ها، جامعیت نسبی بر اساس اصطلاحات مرجع در ایران مدکس (۶۸/۳٪)، مگ ایران (۴۵/۵٪) و اس.آی.دی. (۳۹/۶٪) می‌باشد، در حالی که جامعیت نسبی محاسبه شده بر اساس کلیدواژه‌های نامرجح در ایران مدکس (۷۴/۵٪)، مگ ایران (۴۶/۳٪) و اس.آی.دی. (۴۷٪) است. نتایج در نمودار ۵ قابل مشاهده است. همان‌طور که مشاهده می‌شود، جامعیت نسبی حاصل از نتایج اصطلاحات نامرجح در هر سه پایگاه نسبت به نتایج حاصل از اصطلاحات مرجح بیشتر است.



نمودار ۵. مقایسه جامعیت نسبی پایگاه‌ها بر اساس کلیدواژه‌های مرجح و نامرجح

نتیجه‌گیری

در این پژوهش، تلاش بر این بود تا ضمن مقایسه جامعیت نسبی پایگاه‌های فارسی زبان ایران مدکس، مگ ایران و اس.آی.دی؛ پایگاهی که جامعیت نسبی بیشتری در حوزه دندانپزشکی دارد، مشخص شود. برای محاسبه جامعیت نسبی پایگاه‌ها از روشی که «کلارک و ویلت» در سال ۱۹۹۷ پیشنهاد کرده بودند، استفاده شد. همان‌طور که نتایج نشان داد، جستجو در ایران مدکس، جامعیت نسبی بالاتری نسبت به دو پایگاه دیگر دارد.

از طرفی، استفاده از کلیدواژه‌های اصطلاحنامه پزشکی فارسی، مسائل دیگری را در این پژوهش مورد توجه قرار داد. تصور می‌شد استفاده از اصطلاحات انتخاب شده می‌تواند ملاک مناسبی برای انتخاب کلیدواژه برای انجام جستجو باشد؛ اما در همان ابتدا مشاهده شد که در برخی موارد، حاصل جستجو با کلیدواژه نامرجح، نسبت به نتایج جستجو با کلیدواژه مرجح، جامع تر به نظر می‌رسد. پس از مقایسه میان جامعیت نسبی حاصل از جستجوی کلیدواژه‌های مرجح و نامرجح، مشخص شد در هر سه پایگاه، با استفاده از کلیدواژه‌های نامرجح، جستجوی جامع تری را خواهیم داشت.

گفتنی است، مقایسه‌ای که میان کلیدواژه‌های مرجح و نامرجح انجام شد، تنها محدود به هفت کلیدواژه بود که در این پژوهش کلیدواژه‌های نامرجح آنها نیز مورد جستجو قرار گرفته بود. اینها مواردی بودند که کلیدواژه نامرجح آنها یا در اصطلاحنامه آمده و یا توسط پژوهشگر پیشنهاد شده بود. در حالی که اگر بخواهیم نتایج جستجوی تمامی موارد را بررسی کنیم، با کلیدواژه‌هایی روبرو خواهیم شد که جستجوی آنها هیچ نتیجه‌ای در بر ندارد؛ مانند کلیدواژه «پرستودانتیکس». در هیچ یک از پایگاه‌ها مدرکی با این کلیدواژه در نتایج مشاهده نشد. در ساختار درختی اصطلاحنامه، اصطلاح «پروتز دندان» در زیر «پرستودانتیکس» به عنوان اصطلاح مرتبط ذکر شده است. در حالی که به نظر می‌رسد بهتر بود «پروتز دندان» در رأس قرار می‌گرفت و اصطلاحات «پروتز ثابت» و «پروتز متحرک» که جایشان در اصطلاحنامه کاملاً خالی است، به عنوان موضوعات اخص پیشنهاد می‌شد. همچنین، مشاهده می‌شود که استفاده از املاي «ارتودنسی» (اصطلاح

نامرجح) به جای «ارتدسنسی» (اصطلاح مرجح) در هر سه پایگاه نتایج بیشتری را بازیابی می‌کند.

پیشنهادهای پژوهش

با توجه به موارد فوق به نظر می‌رسد لازم است اصطلاحنامه پزشکی فارسی، به عنوان تنها مرجع موجود برای نمایه‌سازی متون فارسی در حوزه پزشکی، مورد تجدید نظر دوباره قرار گیرد و دقت بیشتری در انتخاب کلیدواژه‌ها به کار رود. همچنین، پایگاه‌هایی که به کار نمایه‌سازی مجله‌های این حوزه مشغول هستند، برای حفظ یکدستی در نمایه‌سازی باید از کلیدواژه‌های کنترل شده استفاده کنند. ضمناً پیشنهاد می‌شود پژوهشهای مشابهی در سایر حوزه‌های پزشکی انجام شود و جامعیت نسبی پایگاه‌ها بر اساس کلیدواژه‌های مرجح و نامرجح مورد مقایسه قرار گیرد.

منابع

- احمدی، حمید (۱۳۸۵). ارزیابی بازیابی اطلاعات. مجله الکترونیکی پژوهشگاه اطلاعات و مدارک علمی ایران (نما)، دوره ۶، شماره ۱. (بازیابی شده از طریق اینترنت) ۱۵ نوامبر ۲۰۰۸. <http://www.irandoc.ac.ir/data/e_j/vol6/ahmadi.htm>
- ارشاد سرابی، رقیه (۱۳۸۰). «بررسی و مقایسه مدلاین‌های رایگان در وب». پایان‌نامه کارشناسی ارشد. دانشگاه علوم پزشکی ایرانی، دانشکده مدیریت و اطلاع‌رسانی پزشکی.
- اکبری، اسماعیل و پرویز شهریاری (۱۳۸۴). «تأثیر اختلاف املائی انگلیسی آمریکایی و بریتانیایی بر میزان جامعیت بازیابی اطلاعات». فصلنامه کتاب. شماره ۶۳: ۵۹-۶۶.

– چشمه سهرابی، مظفر (۱۳۷۸). «تأثیر استفاده از اصطلاحنامه در بانکهای کتابشناختی بر میزان جامعیت، مانعیت و مدت زمان جستجوی اطلاعات بازیابی شده». پایان‌نامه کارشناسی ارشد. دانشگاه تربیت مدرس، دانشکده علوم انسانی.

– رهادوست، فاطمه، مریم کازرانی و دیگران (۱۳۷۷). «اصطلاحنامه پزشکی فارسی: نمایه درختی، نمایه الفبایی، نمایه گردشی، نمایه انگلیسی به فارسی». تهران: سازمان اسناد و کتابخانه ملی جمهوری اسلامی ایران. ۲ج.

– شاکری، صدیقه (۱۳۸۷). میزان جامعیت و مانعیت ابزارهای کاوش فارسی اینترنت در بازیابی اطلاعات در حوزه کتابداری و اطلاع‌رسانی. فصلنامه کتاب. شماره ۷۳: ۱۷۷-۲۰۰. (بازیابی شده از طریق اینترنت) ۲۳ نوامبر ۲۰۰۸

<http://nlai.ir/LinkClick.aspx?fileticket=%2frjDSr9Lelw%3d&tabid=204&mid=312>

- Bar-Ilan, J. (1998) "On the overlap, the precision and estimated recall of search engines. A case study of the query "ERDOS"" *Scientometrics* 42(2):207-228

- O'Grady, T. (2007) "Entrez and BLAST: precision and recall in searches of NCBI databases" *Issues in Science & Technology Librarianship*; Issue 52 viewed 20 Feb 2009, <http://www.istl.org/07-fall/refereed2.html>

- Sarr, M. (2004) "Improving precision and recall using a spellchecker in a search engine. Master's thesis in computer science" M.A. Thesis, Stockholm University, department of numerical analysis and computer science royal institute of technology, Viewed 23 Nov 2008, <[http://www.dsv.su.se/~hercules/papers/SarrNodalida 2003.pdf](http://www.dsv.su.se/~hercules/papers/SarrNodalida%202003.pdf)>

- Shafi, S.M., Rather, R. A. (2005) Precision and recall of five search engine for retrieval of scholarly information in the field of biotechnology. *Webology* 2(2) viewed 27 oct 2008, <http://www.webology.ir/2005/v2n2/a12.htm>

پیوست ۱. فهرست کلید واژه های استفاده شده در پژوهش

کلیدواژه‌هایی که در زیر هریک از ۱۸ مورد اصلی آمده، موارد انتخاب نشده هستند و کلید واژه‌هایی که زیرخط‌دار می‌باشند، توسط پژوهشگر پیشنهاد شده و در اصطلاحنامه (به صورت انتخاب شده یا انتخاب نشده) وجود ندارد..

۱. سیمانی کردن ۱-دن

- سمان

- سمنت

۲. تجهیزات دندان پزشکی ۲/دن

۳. پیمایشهای دندان پزشکی ۳/دن

۴. مدل های دندان ۴/ دن

۵. جفت شدن دندان ها ۵/ دن

- اکلوزن دندان

۶. دندان پزشکی ترمیمی ۶-دن

۷. تشخیص بیماری های دهان ۷- دن

۸. اندودانتیکس ۸- دن

۹. زیبایی در دندان پزشکی ۹/ دن

۱۰. تون بخشی دهان ۱۰- دن

۱۱. ارتدنسی ۱۱- دن

- ارتودانتیکس

- ارتودنسی

۱۲. آسیب شناسی دهان ۱۲- دن

- پاتولوژی دهان

۱۳. دندان پزشکی کودکان ۱۳- دن

- پدودانتیکس

- دندانپزشکی اطفال

۱۴. پرئودانتیکس ۱۴- دن

- پرئودنتیکس

- پرئودنتولوژی

۱۵. دندان پزشکی پیشگیری ۱۵- دن

۱۶. پرستودانتیکس ۱۶- دن

۱۷. تکنولوژی دندان پزشکی ۱۷- دن

۱۸. سفید کردن دندان ۱۸- دن

- بلیچینگ