

بررسی امکان رتبه‌بندی مجدد نتایج جستجو بر مبنای درجه دشواری متن به صورت خودکار: مورد کاوی متون تخصصی حوزه علوم رایانه و فناوری اطلاعات

دکتر شعله ارسطوپور^۱

چکیده

هدف: این پژوهش، امکان‌سنجی استفاده از شاخصهای خوانایی مفهوم پایه و سنتی در بازآرایی و رتبه‌بندی مجدد نتایج جستجو در حوزه علوم رایانه و فناوری اطلاعات است. پیشتر از شاخصهای سنجش خوانایی با هدف درجه‌بندی متون برای آموزش استفاده می‌شد، حال آنکه گرایش به استفاده از این شاخصها در بازیابی اطلاعات، رویکرد جدیدی است که به عنوان یک راهکار جایگزین برای تعیین خودکار سطح مخاطبان یک سند بازیابی شده، مورد استفاده قرار می‌گیرد.

روش‌شناسی: بدین منظور، ابتدا با استفاده از ابزار google trends عبارتهای جستجو با بیشترین فراوانی در سطح اینترنت در حوزه علوم رایانه شناسایی شده و بر این اساس ۲۸ جستجو صورت گرفت. در این مرحله، مدارک بازیابی شده از نظر شاخصهای فلش - دیانی، دامنه مدرک و همگرایی مدرک بررسی و در مرحله بعد نتایج حاصل از این بررسی با نظرهای خوانندگان غیرمتخصص مقایسه شد. بدین منظور، نمونه‌هایی از هریک از این متون برای رتبه‌بندی در اختیار ۳۰ نفر (فاقد تحصیلات آکادمیک در حوزه علوم رایانه) قرار گرفت.

یافته‌ها: در مرحله اول، نتایج بررسیها نشان می‌دهد امکان استفاده از شاخص ترکیبی دامنه - همگرایی مدرک برای متون وجود دارد. یافته‌های مرحله دوم بررسی، از عملکرد بهتر رتبه‌بندی بر مبنای شاخص ترکیبی تشکیل شده حکایت داشت.

کلیدواژه‌ها: شاخص خوانایی فلش - دیانی؛ شاخص خوانایی دامنه مدرک؛ شاخص خوانایی همگرایی مدرک؛ رتبه‌بندی خوانایی در بازیابی.

۱. عضو هیئت علمی مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری (arastoopoor@ricest.ac.ir)

درآمد

بیش از ۸۰ سال است محققان بویژه در حوزه آموزش، توجه خود را به بحث سطح دشواری و سادگی متون معطوف کرده‌اند. پذیره اصلی که در ابتدای امر به شکل‌گیری و رشد این حوزه پژوهشی کمک کرده، باور محققان به این بوده است که می‌توان درجه و سطح دشواری یک متن را از طریق فرمولهایی مشخص نمود و با استفاده از مؤلفه‌های استخراج شده از متن محاسبه و با بهره‌گیری از نتایج آن، متون را سطح‌بندی کرد. نتایج حاصل از این بررسیها در نهایت به فهرستی درجه‌بندی شده از متون می‌انجامید که معلمان با کمک آن متن مناسب برای آموزش را انتخاب می‌کردند. در همین راستا، محققان مختلف در بازه‌های زمانی گوناگون، راهبردها و فرمولهای متفاوتی را ابداع و به آزمون گذارده‌اند. با نگاهی به روند توسعه این حوزه مطالعاتی، می‌توان شاهد تغییر مسیر توجه محققان از شاخصهای واژگانی و نحوی به سمت شاخصهای معنایی بود.

استفاده از شاخصهای سنجش خوانایی و به کار بستن آنها در امر بازیابی اطلاعات، از موضوعاتی است که در سطح بین‌المللی کمتر به آن توجه شده است. تنها در دهه اخیر حرکت‌های جدیدی در عرصه استفاده از این شاخصها مشاهده می‌شود؛ که از این میان می‌توان به تلاش OCLC در استفاده از شاخص سنجه واژگانی^۱ برای جستجوی کتابهای مناسب گروه‌های مختلف سنی کودکان و نوجوانان اشاره کرد (OCLC, 2010- Jul- 15)^۲. از دیگر سو، یکی از زمینه‌های دارای اهمیت در برقراری پیوند میان بازیابی اطلاعات و سنجش خوانایی، استفاده از سنجه‌های خوانایی مفهوم - پایه^۳ است. این سنجه‌ها برای استفاده در بازیابی‌های مرتبط با حوزه‌های موضوعی خاص^۴ مناسب می‌باشند. با نگاهی به متون این حوزه می‌توان این‌گونه برداشت کرد که تاکنون در خارج از ایران در حوزه علوم پزشکی پژوهشهایی صورت گرفته است. این درحالی است که در ایران پژوهشهای صورت

1. Lexile Measure.

2. Your Summer Reading Program Helps Stop Summer Reading Loss [Email] Posted through oclc@oclc.org List-serve.

3. Concept-based Readability Measures.

4. Domain-Specific Information Retrieval.

گرفته در حوزه سنجش خوانایی، از بدو امر تاکنون بیشتر بر سنجش‌های کلاسیک تأکید داشته و به سنجش‌های مفهوم - پایه و در نتیجه بازیابی حوزه وابسته در متون فارسی، توجهی نشده است. این بدان معناست که نه تنها جای چنین پژوهشی در زبان فارسی خالی است، بلکه با توجه به اینکه تاکنون حوزه علوم پزشکی به منزله حوزه‌ای که هم عموم مردم و هم متخصصان در سطوح مختلف با منابع اطلاعاتی آن سر و کار دارند مورد توجه این پژوهشها در خارج از ایران بوده است، سایر حوزه‌های دارای ویژگی مشابه با متون علوم پزشکی (به عنوان مثال، علوم رایانه) در سطح بین‌المللی کمتر مورد توجه واقع شده است. بنابراین، مسئله پژوهش حاضر، مشخص نبودن امکان پیاده‌سازی شاخصهای سنجش خوانایی سنتی (فلش - دیانی) و نیز شاخصهای سنجش خوانایی مفهوم - پایه در بازیابی متون تخصصی فارسی است. به عبارت بهتر، مشخص نیست بازآرایی (رتبه‌بندی مجدد) نتایج بازیابی یک پایگاه در پیوند با پرس و جوهای تخصصی بر مبنای این شاخصها، با رتبه‌بندی کاربران غیر متخصص از نظر درجه دشواری متون تا چه حد همخوانی دارد. حال، با توجه به رسوخ رایانه و فناوری اطلاعات به تمام شئون روزمره زندگی افراد، به جز متخصصان حوزه علوم رایانه، بسیاری از کاربران عادی نیز که فاقد تحصیلات دانشگاهی در این حوزه هستند، به جستجوی اطلاعات در این زمینه دست می‌زنند. به همین دلیل، حوزه علوم رایانه به عنوان بستر اصلی بررسی خوانایی مفهوم - پایه، مد نظر قرار گرفت. رویکرد این پژوهش نه تنها گامی مثبت در جهت پیشبرد پژوهشهای داخلی به منظور همسوسازی دانش در این زمینه با دانش جهانی است، بلکه دستاوردهای آن علاوه بر استفاده‌پذیری بومی، کارایی بین‌المللی در سطح اینترنت - حداقل در پیوند با متون فارسی زبان - خواهد داشت.

مروری بر نوشتارها

در پیوند با سابقه کار در این حوزه، می‌توان مطالعات انجام شده در آن را به سه دسته تقسیم کرد (Zhao & Kan 2010):

۱. **سنجش خوانایی اکتشافی**^۱. پژوهشهای «فوگل و واشبورن» (Vogel & Washburne 1928) و «گری و لیری» (Gray & Leary 1935) از نمونه‌های پیشتاز در این زمینه به‌شمار می‌آیند. اما می‌توان به جرأت گفت، برجسته‌ترین و پرکاربردترین شاخصهای سنجش خوانایی این گروه توسط «فلش - کینکید»^۲، «گانینگ»^۳ و «دیل - چال»^۴ تدوین شده‌اند (دیانی ۱۳۷۹).

۲. **سنجش خوانایی بر پایه یادگیری هدایت شده**^۵. گرایش به این قبیل پژوهشها متضمن دسته‌بندی متون بر مبنای میزان دشواری آنها و سپس اختصاص معیارهایی برای تعیین میزان این دشواری است. از جمله پژوهشهای انجام شده در این حوزه می‌توان به «کالینز - تامپسون و کالان» (Schwarm & Callan 2004)؛ «شارم و استن دورف» (Collins-Thompson & Callan 2004)؛ «هایلمن و همکاران» (Ostendorf 2005)؛ «لروی و همکاران» (Leroy, et al. 2008) و «پیتلر و ننگوا» (Pitler & Nenkova 2008) اشاره کرد.

۳. **سنجش خوانایی حوزه وابسته**^۶. رویکردهای غالب در این گروه از پژوهشها را می‌توان به دو گروه رویکردهای مبتنی بر فهرست واژگان^۷ و رویکردهای هستی‌شناسانه^۸ طبقه‌بندی نمود. رویکرد اول به تعیین دشواری مفهومی متن در یک حوزه با توجه به واژگان موضوعی آن حوزه می‌پردازد. پژوهشهای «کیم و دیگران» (Kim, et al. 2007)؛ و «یا» (Borst, et al. 2008) در این گروه قرار می‌گیرند. در رویکرد دوم، با استفاده از

-
1. Heuristic Readability Measures.
 2. Flesch- Kincaid.
 3. Gunning.
 4. Dale-Chall.
 5. Supervised Learning Approaches.
 6. Domain-Specific Readability Measures.
 7. Wordlist based Approaches.
 8. Ontology based Approaches.

اصطلاحنامه‌ها و یا هستی‌شناسیهای موجود، روابط و فاصله مفهومی میان اصطلاحات بررسی شده و معمولاً با استناد به دو شاخص دامنه مدرک و همگرایی مدرک، خوانایی مفهومی متون در حوزه‌های مختلف تعیین می‌شود. در این رویکرد، می‌توان به پژوهش «یان، سونگ و لی» (Yan, Song & Li 2006) و یا «کار ژائو و کان» (Zhao & Kan 2010) اشاره کرد.

در ایران اگرچه دیانی (۱۳۷۹، ۵) اولین حرکتها در زمینه خوانایی متون را به سال ۱۳۵۵ و انتشار کتاب بسامدی واژه‌ها در نوشته‌های مخصوص عامه مردم می‌داند، پیش از او «غروی» (۱۳۴۴) نوشتاری را در این زمینه منتشر ساخته است. با این حال، اولین پژوهش جدی در این زمینه را می‌توان به سال ۱۳۶۳ توسط شورای کتاب کودک نسبت داد (شعاری نژاد ۱۳۷۴).

این درحالی است که یکی از اصلی‌ترین پژوهشهای مرتبط با سنجش خوانایی نوشته‌های فارسی، مربوط به کار دیانی (۱۳۶۹) است. وی بر اساس مطالعات خود، نه تنها به سنجش خوانایی نوشته‌های فارسی برای سطوح سنی مختلف از کودک تا نوجوان پرداخت، بلکه دو شاخص پرکاربرد سنجش خوانایی (یعنی شاخصهای گاینینگ و فلش) را برای استفاده در زبان فارسی مناسب سازی نمود. از دیگر پژوهشهای این حوزه در ایران می‌توان به «یارمحمدی» (۱۳۷۱)، «طابوسیان و عسکریان» (۱۳۷۴)، «کچرشاوی» (۱۳۷۴) و «دیانی» (۱۳۷۹) اشاره کرد.

به هر روی، شاید بتوان دهه ۱۳۷۰ را دهه شکوفایی انجام پژوهش در زمینه سنجش خوانایی دانست. در دهه ۱۳۸۰ نیز پژوهشهای «مفتون و دقیق» (۱۳۸۰)، «فضل‌اللهی و ملکی توانا» (۱۳۸۹) و «بنی اقبال، بزرگی و رضانی» (۱۳۹۰) از جمله موارد قابل ذکرند. به طور کلی با نگاهی به پژوهشهای انجام شده در این زمینه و بررسی سنجه‌های پیشنهادی، می‌توان به این نتیجه رسید که در سطح جهان حوزه سنجش خوانایی با استناد به دشواری و سادگی واژگان به کار رفته در متن، آغاز و در ادامه با پیشرفتهای حاصل و استفاده از رویکردهای یادگیری هدایت شده، به پیشرفتهای قابل توجهی دست یافته است. با وجود پژوهشهای صورت گرفته، تا پیش از این در ایران با

استفاده از شاخصهای مناسب سازی شده گانینگ و یا فلش، بیشتر بر خوانایی نحوی تمرکز داشته و به همین خاطر بیشتر به متون مناسب کودکان و نوجوانان (متون عمومی) پرداخته‌اند. بنابراین، جای پژوهشی که به صورت اخص از متون عمومی یا مناسب کودکان فاصله بگیرد و شاخصهای سنجش خوانایی را در حوزه‌های تخصصی و یا علمی مورد توجه قرار دهد، در زبان فارسی خالی است.

پرسشهای پژوهش

در این راستا، سه پرسش اساسی مطرح گردید:

۱. چه تفاوتی میان متون تخصصی مورد بررسی در حوزه رایانه از نظر شاخصهای فلش - دیانی، دامنه مدرک و همگرایی مدرک وجود دارد؟
۲. آیا امکان رتبه‌بندی مجدد نتایج جستجو در پایگاه مقالات فارسی مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری^۱ بر مبنای هر سه شاخص فلش - دیانی، دامنه مدرک و همگرایی مدرک در آن واحد، وجود دارد؟
۳. آیا رتبه‌بندی دوباره نتایج بر مبنای این سه شاخص می‌تواند برای کاربران عادی (فاقد تخصص در حوزه رایانه) مفید باشد؟

روش پژوهش و جامعه آماری

پژوهش حاضر از نوع کاربردی بوده و در دو مرحله به روشهای تحلیل محتوا و نیز پیمایشی صورت گرفته است. از روش تحلیل محتوا در مرحله سنجش خوانایی متون با استفاده از شاخصهای سنتی و مفهوم - پایه و از روش پیمایشی برای مقایسه نتایج رتبه‌بندی با نظر مخاطبان غیر متخصص، استفاده شده است. جامعه آماری این پژوهش مقاله‌های فارسی نمایه‌سازی شده در پایگاه مقالات فارسی مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری در حوزه علوم رایانه است. با توجه

۱. این پایگاه صرفاً به عنوان یک بستر برای انجام این بررسی انتخاب شده و امکان انجام چنین بررسی روی هر پایگاه دیگری وجود دارد.

به اینکه هدف اصلی این پژوهش بازآرایی نتایج بازیابی شده از پایگاه بر اساس درجه دشواری بود، به منظور انتخاب مجموعه مدارک پایه، از روش جستجو استفاده شد. ابتدا با استفاده از Google Trends^۱ پرکاربردترین واژگان حوزه رایانه توسط کاربران اینترنتی تعیین شد. در نهایت، با توجه به همپوشانی برخی جستجوها، ۲۸ ترکیب کلی برای جستجو در پایگاه انتخاب شد. هدف اصلی از انتخاب این کلیدواژه‌ها آن بود که اصطلاحهایی از حوزه رایانه و فناوری اطلاعات در پایگاه مقالات فارسی مبنای جستجو قرار گیرند که عملاً بیشتر توسط کاربران در جستجوها استفاده می‌شوند. پس از انجام جستجوها، با انتخاب نتایج صفحه اول بازیابی شده، و در نهایت با حذف موارد تکراری، ۲۱۵ مقاله به منزله بدنه متون برای بررسی انتخاب شد.

مشابه با پژوهشهای دیگر، در این پژوهش نیز سه نمونه ۱۰۰ کلمه‌ای از ابتدا، وسط و انتهای متون استخراج شد. از جمله پژوهشگرانی که در تحقیقات خود به نمونه‌گیری پرداخته‌اند، می‌توان به «دیانی» (۱۳۷۹)، «کرافورد» (Crawford 1984) و «پارکر و هاسبروک» (Parker & Hasbrouck 2001) اشاره نمود.

همچنین با توجه به اینکه اصطلاحنامه INSPEC در پاره‌ای از پایگاه‌های ایرانی برای نمایه‌سازی متون حوزه مهندسی و علوم رایانه استفاده شده است، در پژوهش حاضر نیز به منظور سنجش دامنه مدرک و عمق واژگان از یک سو و نیز محاسبه همگرایی مدرک از سوی دیگر، از این اصطلاحنامه استفاده شد. این اصطلاحنامه پیش‌تر در سال ۱۳۸۴ به فارسی برگردانده شده است. با وجود این، نظر به اینکه ویرایش آخر آن مربوط به سال ۲۰۱۱ بوده است، محقق به منظور استفاده از جدیدترین نسخه اصطلاحنامه، از ویرایش انگلیسی آن استفاده کرد.

در مرحله دوم پژوهش از سی نفر از دانشجویان و دانش‌آموختگانی که در رشته‌هایی غیر از علوم رایانه و یا فناوری اطلاعات مطالعه داشتند، استفاده شد. از نظر حجم نمونه، با نگاهی به متون موجود می‌توان به این نتیجه رسید که قاعده ۳۰ نفر حجم نمونه به منزله

1. <http://www.google.com/trends>.

کمینه حجم نمونه برای یک نمونه نسبتاً بزرگ برای انجام پژوهش‌های مختلف به صورت یک قرارداد عام^۱ در آمده است. «دیانی» (۱۳۸۲) نیز در کتاب خود به این نکته اشاره می‌کند که در بسیاری از تحقیقات تجربی و در بسیاری تحقیقات که متغیر وابسته پایایی نسبتاً کمی دارد، استفاده از نمونه‌هایی با ۳۰ مورد توصیه می‌شود. این درحالی است که در پژوهش‌های مشابه خارجی نیز از نمونه‌هایی با حجم حتی کمتر از ۳۰ نفر نیز استفاده شده است. به عنوان مثال، می‌توان به پژوهش «آرمبروستر و اندرسون» (Armbruster & Anderson 1981) با ۱۷ نفر، و یا پژوهش «پیتلر و ننگوا» (Pitler & Nenkova 2008) با ۳۰ نفر اشاره نمود.

در نهایت، پس از نمونه‌گیری، از پرسشنامه‌ای دو قسمتی برای بررسی نظر مخاطبان در مورد رتبه‌بندی نتایج بازیابی برمبنای شاخصهای مورد مطالعه استفاده شد. قسمت اول به رتبه‌بندی متون اختصاص داشت که از کاربر می‌خواست متون را به ترتیب دشواری رتبه‌بندی کند. در قسمت دوم، طی یک پرسش باز از آنها خواسته می‌شد در مورد عوامل اثرگذار بر مفهوم و یا نامفهوم بودن متون توضیحاتی ارائه دهند. این پرسشنامه به صورت تصادفی در اختیار دانشجویان و دانش‌آموختگانی قرار می‌گرفت که در کتابخانه مرکزی دانشگاه شیراز رفت و آمد داشتند و رشته تحصیلی آنها متفاوت از علوم رایانه و یا فناوری اطلاعات بوده است.

شاخصهای مورد بررسی در پژوهش

در این پژوهش به منظور بررسی و سنجش خوانایی متون، شاخصهای زیر به صورت دستی محاسبه گردید:

شاخص فلش / دیانی (دیانی، ۱۳۷۹)

$$262.835 - 1.015 \left(\frac{\text{تعداد کل واژگان}}{\text{تعداد کل جملات}} \right) - 84.6 \left(\frac{\text{تعداد واجها}}{\text{تعداد کلمات}} \right)$$

دامنه مدرک (Yan, Song & li 2006)

$$Scope(d_i) = e^{-\left(\sum_{i=1}^n depth(c_i)\right)}$$

همگرایی مدرک (Yan, Song & li 2006)

$$Cohesion(d_i) = \frac{\sum_{i,j=1}^n Sim(c_i, c_j)}{\text{Number of Associations}}, \text{ where } n > 1, i < j$$

یافته‌های پژوهش

پیش از پرداختن به پاسخ پرسشهای پژوهش، یادآوری نکاتی چند در پیوند با نحوه پاسخ‌دهی به هر پرسش، توالی پرسشها و نیز نتیجه نهایی ضروری است. پرسش اول، به بحث تفاوت میان متون مختلف از نظر شاخصهای فلش - دیانی، دامنه مدرک و همگرایی مدرک می‌پردازد. پرسش دوم، بحث امکان سنجی رتبه‌بندی نتایج بر مبنای هر سه شاخص فلش - دیانی، دامنه مدرک و همگرایی مدرک، در آن واحد را هدف قرار می‌دهد. از رویکردی دیگر، پرسش سوم به بررسی اثر بخشی هریک از رتبه‌بندیها از نظر خوانندگان می‌پردازد. بدیهی است، چنانچه میان نتایج حاصل از رتبه‌بندی خوانندگان و نتایج حاصل از استفاده از هریک از شاخصهای سنجش خوانایی رابطه وجود داشته باشد، می‌توان شاخصهای مذکور را از نظر سطح‌بندی نتایج جستجو، ارزشمند و مؤثر دانست.

پرسش اول: چه تفاوتی میان متون تخصصی مورد بررسی در حوزه رایانه از نظر

شاخصهای فلش - دیانی، دامنه مدرک و همگرایی مدرک، وجود دارد؟

برای پاسخ به این پرسش، تمامی مقادیر مرتبط با هریک از متون، جداگانه محاسبه شد. سپس برای محاسبه یک مقدار کلی برای هر شاخص در هر مقاله، میانگین مقادیر محاسبه شده برای شاخصهای فلش - دیانی، دامنه مدرک و همگرایی مدرک در هر سه نمونه، محاسبه گردید. در نهایت، برای بررسی تفاوت میان گروه‌های مختلف از نظر هریک از این سه شاخص، از آزمون آنوای یک طرفه استفاده شد. شاخصهای محاسبه شده برای هر مقاله مورد آزمون قرار گرفت که نتایج حاصل از آن، در جدول ۱ به

_____ بررسی امکان رتبه‌بندی مجدد نتایج جستجو ... / ۶۷

نمایش در آمده است. همان‌گونه که در این جدول مشاهده می‌شود، مقدار P در سطح ۰/۰۵ برای هر سه شاخص معنادار است. این مسئله بدان معناست که میان متون هریک از حوزه‌های فرعی ۲۸ گانه مرتبط با علم رایانه تفاوت معناداری وجود دارد. این تفاوت می‌تواند ناشی از درجه دشواری هر متن یا متنهای مرتبط با یک گروه نسبت به سایر متون باشد.

جدول ۱. آماره آنوای یک طرفه برای سه شاخص فلش - دیانی، دامنه مدرک و همگرایی مدرک از نظر حوزه‌های فرعی موضوعی زیر مجموعه علوم رایانه

Sig	F	میانگین توان دوم	درجه آزادی	
۰/۰۰۲	۲/۱۲۴	۲۸۹/۹۷۲ ۱۳۶/۵۴۱	۲۷ ۱۸۷ ۲۱۴	فلش - دیانی بین گروه‌ها درون گروه‌ها جمع
۰/۰۰۰	۳/۷۱۵	۱۳۵۵/۶۵۲ ۳۶۴/۹۰۶	۲۷ ۱۸۷ ۲۱۴	دامنه مدرک بین گروه‌ها درون گروه‌ها جمع
۰/۰۳۳	۱/۶۲۳	۶۳۱/۰۸۸ ۳۸۸/۷۲۸	۲۷ ۱۸۷ ۲۱۴	همگرایی مدرک بین گروه‌ها درون گروه‌ها جمع

برای تشخیص تفاوت میان هریک از گروه‌ها، از آزمون تعقیبی توکی^۱ استفاده شد. در پیوند با هریک از شاخصهای مورد بررسی، مدارک بازیابی شده برای حاصل از جستجو برای عبارتهای «ویروس جدید»، «وب + وب سایت» به بروز این تفاوت شده‌اند. در پیوند با شاخص فلش - دیانی، بانگاهی دقیق‌تر به متون این گروه‌ها می‌توان نتیجه گرفت که در تمام متنهای گروه «ویروس جدید» و بویژه در مواردی که شاخص محاسبه شده برای آنها نشان‌دهنده دشوارتر بودن متن است، از اصل کلمات به انگلیسی نیز استفاده شده است؛ حتی در زمانی که معادل فارسی کلمه نیز آورده شده است. از دیگر ویژگیهای متون این گروه، استفاده بسیار زیاد از آوانگاری کلمات است. آوانگاری کلمات در نهایت به ایجاد واژگانی منجر می‌شود که با رسم‌الخط جدید ممکن است برای خواننده اندکی متفاوت و یا تا حدودی ناآشنا باشد. این در حالی است که در پیوند با شاخصهای دامنه مدرک و همگرایی مدرک، با نگاهی به متون بازیابی شده و بررسی دقیق‌تر آنها می‌توان به این نتیجه رسید که در متون گروه «وب + وب سایت»، از اصطلاحات تخصصی کمتری استفاده شده و جایی که تعداد این اصطلاحات بیشتر است، عمق اصطلاحات کمتر بوده است. به عبارت دیگر واژگان تخصصی این بخش کمتر بوده و آنچه به عنوان واژه‌های تخصصی در متون دیده می‌شود، که یا به عنوان اصطلاح رأس به کار می‌روند و یا به اصطلاح رأس نزدیک‌تر بوده‌اند. به بیان بهتر، در متون این گروه نسبت به سایر گروه‌ها اصطلاحات عمومی‌تری به کار رفته است.

بدیهی است، در یک ساختار درختی چنانچه دو اصطلاح رأس باشند و یا اینکه فاصله کمی با یک اصطلاح رأس داشته باشند، از نظر درجه و یا فاصله به هم نزدیکند. لذا با نگاهی به متنهای بازیابی شده در این گروه، می‌توان دلیل بالا بودن میزان شاخص همگرایی مدرک را به وضوح مشاهده کرد. به عبارت دیگر، در متونی که به توصیف

1. Tukey.

وبسایتها می‌پردازند، عموماً از اصطلاحات جزئی و تخصصی کمتری استفاده شده و همین حداقل از نظر دامنه‌واژگانی تخصصی به کار رفته در متن، باعث انسجام محتوایی بیشتر متن می‌شود. بدیهی است، متنی که به حوزه‌های مختلف موضوعی در یک مجال کوتاه می‌پردازد، درجه انسجام خود را پایین آورده است. به این مورد در پژوهش «یان، سونگ و لی» (Yan, Song & Li 2006) نیز اشاره شده و یافته‌های این پژوهش در جهت یافته‌های آنهاست.

پرسش دوم: آیا امکان رتبه‌بندی مجدد نتایج جستجو در پایگاه مقالات فارسی مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری بر مبنای هر سه شاخص فلش - دیانی، دامنه‌مدرک و همگرایی مدرک در آن واحد وجود دارد؟

پیشنهاد پژوهشگران مختلف برای استفاده از شاخصهای سنجش خوانایی مختلف به صورت همزمان، هنگامی که تمامی شاخصها بر یک پایه باشند، استفاده از جمع است. در این مرحله، با توجه به اینکه تاکنون در این زمینه در زبان فارسی پژوهشی انجام نشده که اثبات کننده امکان انجام چنین کاری حداقل در متون فارسی باشد، به منظور سنجش امکان‌پذیری این کار، وجود رابطه میان شاخصهای محاسبه شده برای متون بررسی شد. بدیهی است، وجود رابطه میان شاخصهای محاسبه شده، نشان‌دهنده وجود هم‌نوایی میان این شاخصهاست و امکان افزودن آنها به یکدیگر را فراهم می‌آورد. بدین منظور، از آزمون وجود رابطه معنادار میان دو متغیر استفاده شد تا شرایط وجود رابطه میان شاخصها به صورت دو به دو بررسی گردد. با نگاهی به جدول ۲ می‌توان نتیجه گرفت که میان دو شاخص دامنه‌مدرک و همگرایی مدرک که هر دو ناظر بر خوانایی محتوایی مدرک هستند، رابطه معنادار و منفی وجود دارد. این نکته بخصوص با توجه به پراکندگی داده‌ها بر قوت این ادعا می‌افزاید، زیرا معمولاً پراکندگی داده‌ها (با استناد به انحراف استاندارد داده‌ها) از به وجود آمدن رابطه قوی جلوگیری می‌کند.

جدول ۲. بررسی وجود رابطه میان شاخصهای محاسبه شده برای مدارک

	فلش دیانی	دامنه مدرک	همگرایی مدرک
فلش دیانی			
آماره پیرسون		- ۰/۰۲۲	۰/۴۰۰
Sig. (دو راهه)	_____	۰/۷۴۴	۰/۹۵
کوواریانس		- ۶/۲۰۱	۱/۱۰۵
دامنه مدرک			
آماره پیرسون	- ۰/۰۲۲		- ۰/۴۹۹
Sig. (دو راهه)	۰/۷۴۴	_____	۰/۰۰۰
کوواریانس	- ۶/۲۰۱		- ۲۲۶/۱۹۰
همگرایی مدرک			
آماره پیرسون	۰/۴۰۰	- ۴/۹۹	
Sig. (دو راهه)	۰/۹۵	۰/۰۰۰	_____
کوواریانس	۱/۱۰۵	- ۲۲۶/۱۹۰	

این نکته بدان معناست که با افزایش درجه دامنه مدرک، از همگرایی مدرک کاسته می‌شود، زیرا هرچه تعداد واژگان تخصصی در متن بیشتر باشد، به واسطه عمیق‌تر شدن واژگان از یک سو و تعداد روابط محتوایی از سوی دیگر، احتمال کاهش میزان همگرایی مفهومی محاسبه شده بیشتر می‌شود. با وجود این همنوایی، استثناهایی نیز وجود دارد. در جایی که متن مورد نظر تنها یک واژه تخصصی داشته باشد، بدیهی است همگرایی متن به حداکثر خواهد رسید. در چنین شرایطی، دامنه مدرک نیز رقم قابل توجهی خواهد بود. گفتنی است، زمانی که دامنه مدرک حداکثر باشد (یعنی ۱۰۰) زمانی است که مدرک هیچ واژه تخصصی را در خود ندارد. بدیهی است، در چنین شرایطی، همگرایی مدرک با توجه به نبود واژه تخصصی از نظر حوزه موضوعی مورد بررسی، صفر خواهد بود.

میان شاخص فلش - دیانی با دو شاخص دیگر، رابطه معناداری مشاهده نمی‌شود. به عبارت دیگر، نتایج این بررسی نشان می‌دهد در رتبه‌بندی نتایج جستجو بهتر است از شاخص فلش دیانی به صورت جداگانه و از شاخصهای دامنه مدرک و همگرایی مدرک در کنار یکدیگر و با هم استفاده شود. بنابراین، پیشنهاد مشخص با توجه به یافته‌های این پژوهش آن است که از شاخص DSDC یا همان دامنه همگرایی مدرک با فرمول زیر به منظور رتبه‌بندی مجدد نتایج استفاده شود.

$$DSDC = \text{همگرایی مدرک} + \text{دامنه مدرک}$$

از دیگر سو، با توجه به اینکه رابطه میان این دو شاخص معنادار است، می‌توان با استفاده از آماره رگرسیون خطی، معادله خط این رابطه را استخراج کرد، تا در شرایطی که امکان محاسبه روابط میان اصطلاحات به کار رفته در متن وجود ندارد، بتوان سطح همگرایی مدرک را برآورد نمود. نتایج ارائه شده در جدول ۳ نشان‌دهنده آن است که می‌توان با اطمینان ۹۹٪ براساس مقدار مشخص دامنه مدرک، به محاسبه همگرایی مدرک پرداخت.

جدول ۳. جدول همبستگی میان شاخصهای دامنه مدرک و همگرایی مدرک

مدل	همبستگی غیر استاندارد		همبستگی استاندارد	t	
	B	خطای استاندارد	Beta		
(ثابت)	۶۸/۵۵۷	۱/۴۳۵		۴۷/۷۷	۰/۰۰۰
دامنه مدرک	- ۰/۴۶۲	۰/۵۵۰	- ۰/۴۹۹	-۸/۴۰۵	۰/۰۰۰

۱. پیش از انجام قطعی این کار و استفاده از آزمون رگرسیون، می‌توان از نمودار توزیع نرمال خطاها استفاده کرد تا بررسی شود که آیا خطاها توزیع نرمالی دارند یا خیر. در این پژوهش، با توجه به نزدیک بودن خط توزیع نرمال خطاها به توزیع واقعی خطاهای استاندارد (نمودار هیستوگرام) و همچنین ناچیز بودن مقدار میانگین می‌توان به این نتیجه رسید که توزیع خطاها نرمال است. به منظور جلوگیری از طولانی شدن نوشتار، از آوردن این نمودار در متن مقاله خودداری شده است.

بنابراین، می‌توان معادله خط رگرسیون دامنه مدرک و همگرایی مدرک را به صورت زیر تدوین نمود:

$$\text{همگرایی مدرک} = ۰/۴۶۲ X + ۶۸/۵۵۷$$

چنان که در جدول ۳ نیز قابل مشاهده است، مقدار ثابتی که باید به معادله اضافه شود، ۶۸/۵۵۷ محاسبه گردیده است. این در حالی است که ضریب محاسبه شده برای دامنه مدرک (X مطرح در معادله) برابر همین جدول ۰/۴۶۲- محاسبه شده است. بنابراین، می‌توان چنین برداشت کرد که با افزایش هریک واحد به میزان دامنه مدرک، ۰/۴۹۹ از میزان همگرایی مدرک کاسته می‌شود. از دیگر سو، با توجه به اینکه در برخی موارد ممکن است مقدار محاسبه شده برای دامنه، همگرایی و یا دامنه - همگرایی مدرک در چندین مدرک بازبایی شده یکسان باشد، می‌توان از شاخص فلش - دیانی برای بازآرایی و رتبه‌بندی آن مدارک در مرحله بعدی استفاده کرد.

پرسش سوم: آیا رتبه‌بندی مجدد نتایج بر مبنای این سه شاخص می‌تواند برای کاربران عادی (فاقد تخصص در حوزه رایانه) مفید باشد؟

به منظور پاسخ به این پرسش، دومین نمونه ۱۰۰ کلمه‌ای استخراج شده از هر مقاله به عنوان متن مرجع برای بررسی خوانایی^۱ در اختیار ۳۰ نفر از دانشجویان و دانش آموختگان دانشگاه شیراز که در کتابخانه مرکزی دانشگاه شیراز حضور داشتند، قرار گرفت. آنها پس از مطالعه این متون، اقدام به رتبه‌بندی آنها از نظر درجه سختی و سادگی کردند.

۱. به منظور انتخاب از بین یکی از سه نمونه ۱۰۰ کلمه‌ای، برای ارائه به خوانندگان جهت رتبه‌بندی، با بررسی آماری داده‌ها، گروهی که پراکندگی در آنها بیشتر بود، به منزله نمونه انتخاب گردید. هدف از این انتخاب آن بود که شانس وجود رابطه اتفاقی میان رتبه‌بندی حاصل از شاخصها و رتبه‌بندی خوانندگان به حداقل رسیده و وجود رابطه نشان از قطعیت بیشتر پاسخ باشد.

جدول ۴. خلاصه آماره گاما بر رتبه‌های شاخص فلش - دیانی، دامنه

مدرک، همگرایی مدرک و دامنه - همگرایی مدرک^۱

Sig	خطای استاندارد	مقدار گاما	
۰/۷۴۱	۰/۱۵۷	- ۰/۰۵۲	رتبه خوانندگان* رتبه فلش
۰/۲۵	۰/۱۵۳	- ۰/۱۸۱	رتبه خوانندگان* دامنه مدرک
۰/۳۲۷	۰/۱۵۰	۰/۱۴۸	رتبه خوانندگان* همگرایی مدرک
۰/۰۳۱	۰/۱۴۶	۰/۵۲۵	رتبه خوانندگان* دامنه - همگرایی مدرک

در نهایت، نتایج حاصل از این بررسی با رتبه کلی محاسبه شده بر مبنای شاخصهای دامنه مدرک، همگرایی مدرک، دامنه - همگرایی مدرک و نیز شاخص فلش - دیانی به صورت دو به دو مقایسه گردید. این مقایسه از طریق تشکیل جدول توافقی و نیز با استفاده از آزمون گاما صورت گرفت (جدول ۴). داده‌های ارائه شده در جدول ۴ بیانگر آن است که در تمام مواردی که در جدول توافقی میان رتبه‌بندی خوانندگان و سایر رتبه‌بندیها تشکیل شده، رابطه معناداری میان رتبه‌بندی خوانندگان و رتبه‌بندیهای حاصل از محاسبه تک‌تک شاخصها وجود ندارد. اما نتیجه جالب در سطر چهارم جدول مشاهده می‌شود. این بررسی نشان می‌دهد رتبه‌بندی انجام شده بر مبنای شاخص ترکیبی دامنه - همگرایی مدرک، رابطه معناداری با رتبه‌بندی خوانندگان دارد. با توجه به عدد ۰/۵۲۵ نیز می‌توان به این نتیجه رسید که این رابطه مثبت است و با بالا رفتن رتبه شاخص دامنه - همگرایی مدرک (یعنی خواناتر شدن متن از نظر محتوایی)، متن به نظر خواننده ساده‌تر می‌آید.

۱. این جدول حاصل ادغام ۴ جدول از خروجی‌های نرم افزار SPSS است.

با توجه به نتایج حاصل از بررسی می‌توان این‌گونه جمع‌بندی نمود که از میان سه شاخص مورد بررسی، هیچ یک به تنهایی نمی‌تواند به بهترین وجه منعکس‌کننده نظریات خوانندگان متون باشد. این درحالی است که شاخص فلش - دیانی یک شاخص سنجش خوانایی سنتی بوده و عملاً تمرکزی بر تخصصی بودن یا نبودن متون ندارد. بنابراین، نمی‌تواند به تنهایی منعکس‌کننده درجه سختی یا سادگی یک متن تخصصی باشد. مخاطبان یک متن تخصصی (ساده یا دشوار) در نهایت به درجه‌ای از توانایی مطالعه رسیده‌اند که معمولاً شاخصهای سنجش خوانایی سنتی که تنها بر سطح اول تحلیل زبانی یعنی تحلیل واژگانی تمرکز دارند، نمی‌توانند آینه تمام‌نمایی برای تعیین درجه دشواری و یا سادگی یک متن تخصصی برای آنها باشند. تاکنون مطالعه‌ای در این زمینه در ایران صورت نگرفته است، اما مطالعات خارجی مانند مطالعات «فنگ» (Feng 2010)، «تامین» (Ta-min 2006)، «سی و کالان» (Si & Callan 2001) و «کالینز تامسون و کالان» (Collins-Thompson & Callan 2001) و بسیاری دیگر نتایج مشابهی را گزارش می‌کنند. شاید به همین دلیل «بنی اقبال، بزرگی و رضانی» (۱۳۹۰) در پژوهش خود تفاوت معناداری میان چکیده‌های ساختار یافته و نیافته بر مبنای شاخص فلش - دیانی مشاهده نکردند. این بدان معناست که در زبان فارسی نیز نمی‌توان تنها با اتکا به شاخصهای سنتی به رتبه‌بندی خوانایی متون تخصصی پرداخت. این در حالی است که نتایج به وضوح نشانگر آن است که شاخص ترکیبی دامنه - همگرایی مدرک که هم به تعداد و عمق واژگان تخصصی توجه دارد و هم به قدرت رابطه میان این واژگان و میزان انسجام و وابستگی آنها، شاخص بهتری در تشخیص درجه دشواری متون تخصصی است. این یافته با یافته‌های «یان، سونگ و لی» (Yan, Song & Li 2006) هماهنگی دارد. بنابراین، می‌توان پس از محاسبه درجه دشواری هر متن بر مبنای این شاخص ترکیبی، گزینه دیگری با عنوان مرتب سازی بر مبنای سطح دشواری مدرک را به صفحه رابط کاربری نتایج جستجو افزود.

خوانندگان در پاسخ به پرسش باز مطرح در زمینه عوامل اثرگذار بر دشواری یک متن تخصصی، به نکات جالب توجهی اشاره کردند. از جمله این موارد، بحث استفاده

از واژگان تخصصی به صورت‌های مختلف در متون تخصصی است. اکثر این واژگان تخصصی از زبانهای خارجی (بویژه انگلیسی) وارد زبان فارسی شده و به واسطه تحول سریع در رشته، فرصت معادل‌سازی مناسب برای آنها وجود نداشته است؛ و یا گاه به واسطه سرعت شیوع استفاده از یک فناوری در سطح جامعه و کاربرد وسیع واژه خارجی توسط کاربران، استفاده از واژه معادل سازی شده در یک متن از دید خوانندگان آن، به افزایش پیچیدگی آن متن منجر گشته است. یکی دیگر از مشکلات مورد توجه خوانندگان، استفاده از آوانگاری واژگان خارجی است. همین تبدیل از یک فرم نگارشی به فرم آوانگاری برای برخی از خوانندگان به دشوار شدن یک متن می‌انجامد.

با نگاهی به پاسخها، گرایش خوانندگان عادی متون تخصصی در حوزه علوم رایانه یا به استفاده از عین کلمات به انگلیسی با همان حروف انگلیسی و یا استفاده از برابر نهادهای فارسی بوده و استقبال از عین کلمه به صورت آوانگاری شده در بین خوانندگان بسیار کم بود. به عنوان نمونه، یکی از خوانندگان چنین استدلال می‌کند: «استفاده از واژگان انگلیسی با همان حروف بدون تغییر، باعث می‌شود هنگام کار عملی با کلمه‌ها آشنا بوده و کاربر دچار سردرگمی نشود». خواننده دیگری در همین رابطه بیان می‌دارد: «هر علم، واژگان مختص به خود را دارد که بالطبع اولین زبانی که علم در آن قرار دارد، باید مرجع قرار گیرد».

از این یافته‌ها می‌توان در دسته‌بندیهای ثانویه به منظور تشخیص درجه دشواری یک متن برای خوانندگان عادی براساس نوع واژگان به کار رفته در متن استفاده کرد. اما به نظر می‌رسد ارائه نظر قطعی در این زمینه، به بررسیهای بیشتر به منظور رسیدن به قواعدی مستدل وابسته است.

بحث و نتیجه‌گیری

با نگاهی به یافته‌های این پژوهش در پیوند با پرسشهای اول و دوم، می‌توان به این نتیجه رسید که امکان استفاده از شاخصهای پیش‌گفته در متون فارسی وجود دارد. در

پیوند با شاخص فلش، دیانی (۱۳۷۹) سالها پیش امکان استفاده از این شاخص را در صورت مناسب سازی شده برای ویژگیهای زبان فارسی به اثبات رسانده بود. به همین دلیل، شاخص مناسب سازی شده فلش - دیانی را می‌توان بدون هیچ مشکلی در مورد نوشتارهای فارسی عمومی به کار برد. این در حالی است که بر اساس بررسیهای انجام شده، استفاده از دو شاخص «دامنه» و «همگرایی مدرک»، امکان‌پذیر است و با توجه به آنچه در پیوند با پرسش دوم مطرح گردید، قابلیت ادغام دو شاخص دامنه و همگرایی مدرک در قالب یک شاخص ترکیبی فراهم است.

در مرحله دوم پژوهش، رتبه‌بندی خوانندگان عادی با رتبه‌بندی انجام شده بر مبنای شاخصهای مورد نظر مقایسه گردید. نتایج این مقایسه حاکی از آن است که تنها میان رتبه‌بندی خوانندگان عادی با رتبه‌بندی حاصل از شاخص ترکیبی دامنه - همگرایی مدرک، رابطه‌ای معنادار وجود دارد. از آنجا که شاخص دامنه مدرک بر تعداد و عمق واژگان تخصصی تمرکز دارد و شاخص همگرایی مدرک بر روابط مفهومی میان این واژگان و طول این روابط متمرکز است، چنین می‌توان برداشت کرد که صرف تعداد و عمق واژگان تخصصی و یا صرف نزدیکی روابط میان اصطلاحات تخصصی، هیچ یک به تنهایی نمی‌تواند به عنوان شاخص اصلی رتبه‌بندی برای تعیین سطح مخاطبان یک مدرک، ملاک عمل قرار گیرند. این در حالی است که شاخص ترکیبی دامنه - همگرایی مدرک که ناظر به هر دو زمینه عمق واژگان و روابط میان آنهاست، به مراتب به رتبه‌بندی خوانندگان عادی نزدیکتر است. از سوی دیگر، مشابه با پژوهشهای انجام شده در خارج از ایران، نتایج این پژوهش نیز حاکی از آن است که نمی‌توان تنها به شاخص سنتی فلش - دیانی برای رتبه‌بندی مدارک استناد کرد. البته، همان‌گونه که پیشتر نیز بیان شد، این یافته ناقض کارایی فرمول پیش گفته در سنجش خوانایی نوشته‌های فارسی بدون در نظر گرفتن تخصص موضوعی نیست. نکته‌ای که باید بر آن تأکید کرد، آن است که وقتی بحث خوانایی متون تخصصی مطرح می‌گردد، مفهوم بودن متن برای خواننده آن بیش از آنکه متوجه شاخصهای مبتنی بر ویژگیهای واژگانی و نحوی باشد، متوجه ویژگیهای محتوایی است.

بنابراین، با توجه به یافته‌های این پژوهش می‌توان گفت نه تنها امکان استفاده از شاخص ترکیبی دامنه - همگرایی مدرک برای بازآرایی نتایج براساس درجه دشواری وجود دارد، بلکه رتبه‌بندی نتایج بازیابی شده بر اساس این شاخص، با نظر خوانندگان بالقوه هماهنگی بیشتری دارد. این ویژگی بخصوص در مواردی که مدارک بازیابی شده از نظر ربط در مرتبه یکسانی است، کاربرد می‌یابد، زیرا کاربر در مواجهه با مدارکی که از نظر ربط مشابه هستند، می‌تواند مدارکی را انتخاب کند که از نظر سطح تخصصی، مناسب با توان علمی وی باشد. مسلماً چنین راهبردی کمک شایانی به متخصصان نیز خواهد کرد، زیرا آنها نیز در انتخاب خود می‌توانند مدارکی را انتخاب و ذخیره کنند که در سطح تخصص آنها باشد. در چنین شرایطی، در زمان اختصاص یافته از سوی متخصصان نیز صرفه جویی خواهد شد.

منابع

- بنی‌اقبال، ناهید؛ اشرف‌السادات بزرگی و ابوذر رضانی (۱۳۹۰). بررسی چکیده‌های موجود در فصلنامه کتابداری کتابخانه و مرکز اسناد دانشگاه تهران براساس الگوی چکیده ساختاریافته. *فصلنامه دانش‌شناسی (علوم کتابداری و اطلاع‌رسانی و فناوری اطلاعات)*، جلد ۳ (۱۲): ۱-۱۵.
- دیانی، محمدحسین (۱۳۶۹). معیاری برای سطح خوانایی نوشته‌های فارسی. *مجله علوم اجتماعی و انسانی دانشگاه شیراز*. شماره ۲.
- _____ (۱۳۷۹). سنجش خوانایی نوشته‌های فارسی. مشهد: کتابخانه رایانه‌ای.
- _____ (۱۳۸۲). گلوگاه‌های پژوهش در علوم اجتماعی. مشهد: کتابخانه رایانه‌ای.
- شعاری‌نژاد، علی‌اکبر (۱۳۷۴). ادبیات کودکان. تهران: اطلاعات.
- طابوسیان، جلیله و سودابه عسکریان (۱۳۷۴). ارزشیابی کتب ادبیات فارسی دوره راهنمایی تحصیلی از نظر سطح خوانایی متن بر اساس فرمول فلش. پایان‌نامه کارشناسی، دانشگاه پیام نور، اهواز.
- غروی، مهدی (۱۳۴۴). نویسندگی برای نوسودان. *مجله آموزش و پرورش (تعلیم و تربیت)*. شماره ۳۵۵: ۳۸-۴۹.
- فضل‌اللهی، سیف‌الله و منصوره ملکی‌توانا (۱۳۸۹). روش‌شناسی تحلیل محتوا با تاکید بر تکنیک‌های خوانایی سنجی و تعیین ضریب درگیری متون. *پژوهش*، جلد ۲ (۱): ۷۱-۹۴.

- کچرشاوی، مجیده (۱۳۷۴). ارزشیابی کتابهای علوم اجتماعی دوره ابتدایی و راهنمایی بر اساس سطح خوانایی متن. پایان‌نامه کارشناسی، دانشگاه پیام نور، اهواز.
- مفتون، پرویز و مهتا دقیق (۱۳۸۰). معیار تعیین میزان خوانایی ترجمه‌های فارسی متون انگلیسی. *پژوهشنامه علوم انسانی*، شماره ۲۹: ۶۱ - ۷۹.
- یارمحمدی، مجتبی (۱۳۷۱). *تحقیقی پیرامون سطح خوانایی متون فارسی بر اساس فرمول پیشنهادی دکتر محمدحسین دیانی*. تهران: گروه تحقیق و ارزیابی مدیریت آموزش مداوم.
- Armbruster, B. & T. Anderson. (1981). Content Area Textbooks (Reading Education Report No. 23). Urbana, IL: Center for the Study of Reading.
- Borst, A. et.al. (2008). Lexically Based Distinction of Readability Levels of Health Documents, *Acta Informatica Medica* 16(2): 72 - 75.
- Collins-Thompson, K. & J. Callan. (2004). A language modeling approach to predicting reading difficulty. In Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2004).
- Crawford, A. N. (1984). *A Spanish Language Fry-Type Readability Procedure: Elementary Level* (Vol. 7 No. 3). Los Angeles: California State University, Evaluation, Dissemination and Assessment Center.
- Feng, L. (2010). *Automatic Readability Assessment*. New York: City University of New York.
- Gray, W & B. Leary. (1935). *What Makes a Book Readable?* Chicago: Chicago Press.
- Heilman, M. et.al. (2007). Combining Lexical and Grammatical Features to Improve Readability Measures for First and Second Language Texts. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics*.
- Kim, H. et.al. (2007). Beyond Surface Characteristics: A New Health Text- Specific Readability Measurement. In *AMIA Annual Symposium Proceedings*. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2655856/> (Accessed 2012-03-13)
- Leroy, G.; et.al. (2008). A Balanced Approach to Health Information Evaluation: A Vocabulary-Based Naïve-Bayes Classifier and Readability Formulas. *Journal of the American for Information Science and Technology*, Vol 59(9): 1409 - 1419.
- Parker, R. I., & J. E. Hasbrouck. (2001). Spanish readability formulas for elementary-level texts: A validation study. *Reading and Writing Quarterly*, 17, 307-322.
- Pitler, E. & A. Nenkova. (2008). Revisiting Readability: A Unified Framework for Predicting text Quality. In *EMNLP*, P. 186 - 195.
- Schwarm, S. & M. Ostendorf. (2005). Reading Level Assessment Using Support Vector Machines and Statistical Language Models. In *ACL's 05*. P. 523 - 530.
- Si, L. & J. Callan. (2001). A Statistical Model For Scientific Readability. In *Proceedings of the 10th International Conference on Information and Knowledge Management*.
- Ta-Min, R. (2006). *Comprehension of Online Cancer Information: A Propositional Assessment of Readability, Inferences, and Coherence*. Ontario: University of Waterloo.

- Vogel, M. & C. Washburne. (1928). An Objective method of Determining Grade Placement of Children's Reading Material. The Elementary School Journal, No. 28: 373 – 381.
- Yan, X. ; D. Song, & X. Li. (2006). Concept-based Document Readability in Domain Specific Information Retrieval. In CIKM ' 06, P. 540 – 549.
- Zhao, J. & M. Kan. (2010). Domain Specific Iterative Readability Computation. In JCDL '010. P. 250 – 214.

