

A Critique of the Book: *Natural Language Processing and Information Retrieval*

Mortaza Kokabi*

Abstract

The book entitled “*Natural Language Processing and Information Retrieval*” is a compiled and translated book, without any specific author, that has been translated by Jaafar Mehrad and Maryam Naseri. If we assume that information retrieval is the heart of the field of knowledge and information science, then we can claim that if it is done in user’s natural language, the language by which she/he lives and seeks information, the result will be much more useful for her/him. Therefore, a book on *Natural Language Processing and Information Retrieval* will be of great value for the information field and profession. Furthermore, the introduction of a course entitled “Natural Language Processing” in PhD curriculum with a minor of “Information Retrieval” necessitates the existence of such a book due to the shortage (and perhaps lack of) such a source for this course. However, there are some problems with the book that will be discussed in some detail in the section on the critique and evaluation of book’s content. Some other discussions will also be presented on the citation and referencing problems in the same section. Using some tables, the relationship among the terms translated in the book, vocabulary, and index will also be discussed.

Keywords: Natural Language Processing, Information Retrieval, Spanish Language, English Language, Arabic Language, Farsi Language

* Prof. Emeritus, Department of Knowledge and Information Science, Shaheed Chamran University, Kokabi80@yahoo.com

Date received: 2020-04-21, Date of acceptance: 2020-10-31

Copyright © 2010, IHCS (Institute for Humanities and Cultural Studies). This is an Open Access article. This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.



پروہشگاہ علوم انسانی و مطالعات فرہنگی
پرتال جامع علوم انسانی

نقدی بر کتاب پردازش زبان طبیعی و بازیابی اطلاعات

مرتضی کویکی*

چکیده

کتاب «پردازش زبان طبیعی و بازیابی اطلاعات» کتابی است گردآورده و ترجمه شده که نویسنده مشخصی ندارد و ترجمه جعفر مهران و مریم ناصری است. اگر بپذیریم که بازیابی، قلب رشته علم اطلاعات و دانش‌شناسی است آن گاه می‌توانیم بگوییم که اگر این بازیابی با زبان طبیعی انجام شود، نتیجه برای او بسیار سودمندتر خواهد بود. پس کتابی که پیرامون «پردازش زبان طبیعی و بازیابی اطلاعات» باشد می‌تواند برای رشته و حرفه از ارزش بالایی برخوردار باشد. افزون بر این، در گرایش «بازیابی اطلاعات و دانش» دوره دکتری علم اطلاعات و دانش‌شناسی، از سال ۱۳۹۱ درسی تخصصی الزامی با عنوان «پردازش زبان طبیعی» ارائه شده که با توجه به کمبود (و شاید هم نبود) منابعی برای این درس، وجود کتابی با محتوای پردازش زبان طبیعی می‌تواند فی‌نفسه سودمند باشد. اما در کتاب، ایرادهایی وجود دارد که در تحلیل و ارزیابی محتوایی اثر با تفصیل بیشتری پیرامون آن‌ها بحث خواهد شد. درباره رعایت اصول علمی ارجاع‌دهی در درون متن و کتاب‌نامه‌ها که در پایان هر فصل آمده‌اند نیز در تحلیل و ارزیابی محتوایی اثر مطالبی ارائه خواهند شد. رابطه میان اصطلاحات ترجمه شده در متن، واژه‌نامه و نمایه نیز در کتاب مورد نقد، با استفاده از جدول‌هایی بررسی خواهد شد.

کلیدواژه‌ها: پردازش زبان طبیعی، بازیابی اطلاعات، زبان اسپانیایی، زبان انگلیسی، زبان عربی، زبان فارسی

* استاد بازنشسته علم اطلاعات و دانش‌شناسی، دانشگاه شهید چمران اهواز، Kokabi80@yahoo.com

تاریخ دریافت: ۱۳۹۹/۰۲/۰۲، تاریخ پذیرش: ۱۳۹۹/۰۸/۱۰

۱. مقدمه

عبارت «پردازش زبان طبیعی» به‌عنوان درسی از درس‌های گرایش «بازیابی اطلاعات» دوره دکترای علم اطلاعات و دانش‌شناسی، شاید نخستین بار به‌طور رسمی در مصاحبه جعفر مهرداد، سرپرست پیشین کمیته برنامه‌ریزی علم اطلاعات و دانش‌شناسی وزارت علوم، تحقیقات و فناوری در مصاحبه با خبرگزاری لیژنا اعلام شد. بخش‌هایی از این مصاحبه عیناً در زیر آورده شده است:

جعفر مهرداد، سرپرست کمیته برنامه‌ریزی علم اطلاعات و دانش‌شناسی وزارت علوم در گفتگو با لیژنا با اعلام خبر تصویب اولین برنامه دوره دکتری علم اطلاعات و دانش‌شناسی اظهار داشت: از سال ۱۳۸۶ تا کنون، برنامه‌های مختلفی در مقاطع مختلف تحصیلی توسط این کمیته برنامه‌ریزی و به تصویب رسیده که مصوبات کمیته رسماً توسط وزارت علوم، تحقیقات و فناوری به دانشگاه‌ها ابلاغ شده است.

وی افزود: در جلسه ۱۳۹۱/۱۲/۱۵ بعد از چندین ماه برنامه‌ریزی درباره گرایش بازیابی اطلاعات و دانش در مقطع دکتری، لاجرم این گرایش که نخستین برنامه دوره دکتری علم اطلاعات و دانش‌شناسی است به تصویب رسید.

به گفته مهرداد، تعداد درس‌های پایه این گرایش چهار درس به شرح زیر می‌باشد:

۱- فلسفه علم

۲- نظریه‌های اطلاعات و دانش

۳- طبقه‌بندی و توسعه علوم

۴- سبیرنتیک و ارتباطات

وی تصریح کرد: برای این گرایش ۶ درس تخصصی الزامی نیز به تعداد ۱۲ واحد منظور شده است:

۱- پردازش زبان طبیعی

۲- حوب معنایی و هستی‌شناسی

۳- تحلیل متن و دانش کاوی

۴- معماری اطلاعات

۵- ارزیابی نظام‌های بازیابی اطلاعات و دانش

۶ سمینار مسائل ویژه در بازیابی اطلاعات و دانش (تصویب اولین برنامه دکتری علم اطلاعات و دانش‌شناسی، ۱۳۹۱).

همان گونه که مشاهده می‌شود درس «پردازش زبان طبیعی» نخستین درس از ۶ درس تخصصی الزامی در گرایش «بازیابی اطلاعات و دانش» از دوره دکتری علم اطلاعات و دانش‌شناسی است. قرار دادن دو عبارت «پردازش زبان طبیعی» در کنار عبارت «بازیابی اطلاعات و دانش» عبارتی را درست می‌کند که بی‌شبهت به عنوان کتاب «پردازش زبان طبیعی و بازیابی اطلاعات» نیست. در این باره در جای دیگری از این نوشته بحث خواهد شد.

کتاب «پردازش زبان طبیعی و بازیابی اطلاعات» کتابی است گردآورده و ترجمه شده که نویسنده مشخصی ندارد و ترجمه جعفر مهران و مریم ناصری^۱ است. البته نویسنده (نویسندگان) هر فصل مشخص است (هستند) و نشانی‌های پست الکترونیکی برخی از آنان نیز داده شده است. در مورد نویسندگان فصل‌ها توضیح چندانی داده نشده و اطلاعات اندکی را می‌توان از نشانی پست الکترونیکی برخی از آنان به دست آورد. جعفر مهران یکی از قدیمی‌ترین استادان رشته علم اطلاعات و دانش‌شناسی در کشور و صاحب انتشارات بسیاری است. او در سمت‌های اجرایی رده بالایی نیز فعالیت داشته و در کل می‌توان او را جزء نامورترین استادان رشته به‌شمار آورد.

کتاب دارای هشت فصل به شرح زیر است:

فصل اول: پردازش زبان طبیعی در بازیابی اطلاعات

فصل دوم: پردازش زبان طبیعی در بازیابی اطلاعات متنی

فصل سوم: نقش پردازش زبان طبیعی در نظام بازیابی اطلاعات هوشمند

فصل چهارم: بازیابی اطلاعات با دقت بیشتر با استفاده از فنون پردازش زبان طبیعی

فصل پنجم: الگوریتم ریشه‌سازی برای زبان فارسی

فصل ششم: جستجوی فارسی و فناوری‌های نمایش

فصل هفتم: ریشه‌سازی و ان-گرام‌ها در اسپانیولی و تاثیر [تأثیر] آنها در بازیابی

اطلاعات

فصل هشتم: پردازش زبان طبیعی عربی برای بازیابی اطلاعات

کتاب، افزون بر این هشت فصل، یک واژه‌نامه انگلیسی-فارسی و یک نمایه نیز دارد.

اما در کتاب، ایرادهایی وجود دارد که در تحلیل و ارزیابی محتوایی اثر با تفصیل بیشتری پیرامون آن‌ها بحث خواهد شد. درباره رعایت اصول علمی ارجاع‌دهی در درون متن و کتاب‌نامه‌ها که در پایان هر فصل آمده‌اند نیز در تحلیل و ارزیابی محتوایی اثر مطالبی ارائه خواهند شد. رابطه میان اصطلاحات ترجمه شده در متن، واژه‌نامه و نمایه نیز در کتاب مورد نقد، با استفاده از جدول‌هایی بررسی خواهد شد.

۲. روش‌شناسی نقد حاضر

روش نقد حاضر، آمیزه‌ای از روش‌های نقد ساختاری و محتوایی است. در نقد ساختاری، مقدمه، متن، نتیجه‌گیری، تقسیم‌بندی اجزای متن، و شیوه سنددهی و ارجاع نقد و ارزشیابی می‌شوند. در نقد محتوایی، محتوا از منظرهای علمی، اخلاقی، دینی، تاریخی و غیره نقد می‌شود. بدیهی است کتابی را که پیرامون پردازش زبان طبیعی و بازیابی اطلاعات است نمی‌توان از منظرهای اخلاقی و دینی نقد کرد، اما آن چه که در نقد حاضر بیشتر مورد توجه قرار می‌گیرد نقد علمی محتواست.

۳. بدنه اصلی مقاله

۱.۳ تحلیل و ارزیابی ظاهری اثر

طرح جلد کتاب که طراح آن کریم فلاح است دارای رنگ‌های گرمی (نارنجی و قرمز) است اما تصویر روی آن معلوم نیست که نشان دهنده چیست (حد اقل نگارنده این سطرها از آن چیزی نفهمیده است). صحافی، قطع، نوع و اندازه قلم، و صفحه‌آرایی کتاب مناسب است.

۲.۳ تحلیل و ارزیابی محتوایی اثر

«پردازش زبان طبیعی و بازیابی اطلاعات» از جمله مفاهیمی است که در رشته علم اطلاعات و دانش‌شناسی از اهمیت زیادی برخوردار است. اگر بپذیریم که بازیابی، قلب رشته علم اطلاعات و دانش‌شناسی است آن گاه می‌توانیم بگوییم که اگر این بازیابی با زبان طبیعی انجام شود، زبانی که کاربر با آن زندگی، و طبعاً اطلاعات را جستجو می‌کند نتیجه برای او بسیار سودمندتر خواهد بود. پس کتابی که پیرامون «پردازش زبان طبیعی و بازیابی

اطلاعات» باشد می‌تواند برای رشته و حرفه از ارزش بالایی برخوردار باشد. افزون بر این، در بالا نیز به این نکته اشاره شد که در گرایش «بازیابی اطلاعات و دانش» دوره دکترای علم اطلاعات و دانش‌شناسی، از سال ۱۳۹۱ درسی تخصصی الزامی با عنوان «پردازش زبان طبیعی» ارائه شده که با توجه به کمبود (و شاید هم نبود) منابعی برای این درس، وجود کتابی با محتوای پردازش زبان طبیعی می‌تواند فی‌نفسه سودمند باشد.

کتاب «پردازش زبان طبیعی و بازیابی اطلاعات» بی‌گمان با این دیدگاه ترجمه شده، به‌ویژه این که عنوان کتاب، ترکیبی از عنوان‌های درس تخصصی الزامی «پردازش زبان طبیعی» و گرایش «بازیابی اطلاعات و دانش»، و گزینش این عنوان احتمالاً برای رساندن این مفهوم است که این کتاب، می‌تواند به‌عنوان متنی درسی برای این درس انتخاب شود. یکی از مترجمان، جعفر مهرداد که از استادان قدیمی رشته است می‌باشد که از نیاز رشته و حرفه به این کتاب، نیک آگاه، و به همین دلیل در ترجمه این اثر مشارکت داشته است. به این ترتیب، اهمیت وجود چنین کتابی بر کسی پوشیده نیست. صرف ترجمه و انتشار چنین کتابی خود، یکی از نکات مثبت کتاب است.

اما از آن جایی که هدف نقد، بیان اشکالات در کنار محاسن اثر است آن چه که در زیر می‌آید نگاهی به محتوای فصل‌های گردآورده و ترجمه شده است. در پیوند با گردآوری و ترجمه، یکی از نکاتی که در کتاب به چشم می‌خورد استفاده از واژه «ترجمه» تنها در روی جلد و صفحه‌عنوان کتاب است در حالی که در فهرست‌نویسی پیش از انتشار برای مترجمان از واژگان «گردآورنده، مترجم» استفاده و در یک یادداشت نیز به گردآوری اشاره شده است. بررسی فصل‌های کتاب نیز نشان می‌دهد که این فصل‌ها از منبع یکسانی گرفته نشده‌اند (یا حداقل چنین به نظر می‌رسد). در فیپای کتاب، یادداشت عنوان به زبان اصلی وجود ندارد و در خود کتاب نیز، نام(های) نویسنده (نویسندگان) ذکر نشده‌اند. مشخصات کتابشناختی فصل‌های کتاب نیز در هیچ جایی داده نشده‌اند. به این ترتیب، نخستین پرسشی که به ذهن خواننده خطور می‌کند این است که فصل‌های کتاب، چگونه و با چه معیاری (یا معیارهایی) انتخاب شده‌اند و ترتیب ارائه این فصل‌ها چیست. به این پرسش در هیچ جای کتاب پاسخ داده نشده است. حتی در مقدمه کتاب که به قلم استاد قدیمی نوشته شده و نیازمند ویراستاری ادبی اندکی است هیچ اشاره‌ای به این نکته نشده است. تنها در صفحه سوم از مقدمه چنین نوشته شده است: «مطالبی که در فصول مختلف این کتاب در پیش رو دارید گزیده‌ای است آگاهانه از مطالعاتی که در مراکز علمی جهان در حیطه زبان [زبان‌های]

فارسی، اسپانیولی و انگلیسی صورت گرفته است.» اما پیرامون این که این گزینش آگاهانه چگونه انجام شده توضیحی داده نشده، ضمن این که در همین جمله، زبان عربی نیز از قلم افتاده است. در س. ۲ از ص. ۲ کتاب، عبارت «این فصل» وجود دارد. معلوم نیست که منظور از این عبارت چیست؛ عبارتی که در همین صفحه هم تکرار شده است. در س. ۱ از ص. ۴ نیز، عبارت «... مثال‌های ۴-۱ پیش گفته را مدیریت می‌کند.» دیده می‌شود. اما این مثال‌ها در این فصل دیده نمی‌شوند. مجموعه این عوامل نشان می‌دهند که کتاب، گردآورده از چند منبع است که به دلایلی معرفی نشده‌اند. در س. اول از ص. ۳۲ (فصل سوم) نیز واژه «فصل» دیده می‌شود. در همین فصل، در عنوان‌های برخی شکل‌ها، به شکل‌ها و جدول‌هایی ارجاع داده شده که در این فصل دیده نمی‌شوند. در ص. ۵۴ نیز واژه «فصل» دیده می‌شود. علت (علل) و معیار(های) این محذوفات، اگر وجود دارند چه بوده‌اند؟ همان گونه که در بالا اشاره شد ترتیب فصل‌ها نیز جای بحث دارد: مثلاً معلوم نیست که چرا مطالب مربوط به زبان‌های فارسی و عربی در آخر آمده‌اند در حالی که مطالب مربوط به زبان اسپانیایی (که معلوم نیست چرا اصلاً انتخاب شده) میان این دو آمده است. ابهام پیرامون محتوای کتاب به همین جا ختم نمی‌شود: ۱۴ منبع از ۳۸ منابع ارجاعات فصل اول (۳۶/۸ درصد) ظاهراً در متن، استفاده نشده‌اند. دلیل این امر چیست؟ آیا در متن اصلی، پیش از ترجمه، حذف‌هایی انجام شده است؟ اگر آری، ملاک محذوفات چه بوده است؟ اگر نه، به چه دلیلی این منابع در ارجاعات وجود دارند اما در متن، استفاده نشده‌اند؟ همانند این اشکال در برخی فصل‌های دیگر نیز دیده می‌شود: ۴ منبع از ۱۴ منبع فصل دوم (۲۸/۵۷ درصد)، ۴ منبع از ارجاعات فصل سوم، ۵ منبع از ارجاعات فصل هفتم و ۳ منبع از ارجاعات فصل هشتم در متن استفاده نشده‌اند.

در بالا اشاره شده که مقدمه کتاب به ویراستاری ادبی اندکی نیاز دارد. نمونه‌ای از این نیاز به ویراستاری در جمله زیر از مقدمه دیده می‌شود: "... اساس نمایه‌سازی افزایش دقت (یعنی نسبت مدرک [مدارک] بازیابی شده [شده‌ای] که مربوط می‌باشند) و [به] بازیافت (یعنی، نسبت مدارک مربوط [مربوطی] که بازیابی می‌شوند) می‌باشد." در پاراگراف سوم از صفحه ه مقدمه جمله‌ای وجود دارد که معنای آن چندان رسا نیست: "به خاطر توجه به علایق خوانندگان این اثر از اصطلاح قدیمی‌تر به عنوان بازیابی مدرک و از اصطلاح جدیدتر به عنوان بازیابی متن به طور مترادف استفاده شده است. هدف هر دو بازیابی متون است: از یک پاراگراف گرفته تا متن یک کتاب." یا در سطر ۲۰ از همین صفحه ظاهراً

بازیابی داده‌ها با بازیابی دانش یکی گرفته شده است در حالی که این دو متفاوتند. در سطرهای ۱۴-۱۸ ص. و مقدمه نیز این پاراگراف دیده می‌شود:

توصیفگرها که ممکن است غیر مستقیم باشند چون متون کامل مدارک که به طور فزاینده‌[ای] به صورت روی خط در دسترس می‌باشند، موضوع توصیفگرها در مقایسه با سالهای گذشته بیشتر تحت کنترل روش‌های نمایه‌سازی است. با وجود این، همیشه نمی‌توان توصیفگرها را به طور کامل تهیه دید. نویسنده یک مدرک همیشه گفته‌های فراوانی دارد که در مدرک به طور کامل ارائه نمی‌دهد.

مفهوم این پاراگراف مبهم و جمله اول آن ناقص است. جمله پس از این پاراگراف هم باید اصلاح شود: آیا «تکالیف ناسازگار» مفهوم درستی است؟ این اشکالات می‌باید در ویراست‌های بعدی کتاب برطرف شوند.

این نیاز به ویراستاری ادبی در جاهای دیگری از کتاب هم دیده می‌شود و تعداد این موارد کم هم نیست. این موارد شامل اصطلاحاتی هستند که درست معادل‌یابی نشده یا جملاتی هستند که یا درست ترجمه نشده یا تحت‌اللفظی ترجمه شده‌اند. یکی از این موارد، عبارت «نظام بازیابی اطلاعات هوشمند» است که از آن جایی که مفهومی به‌نام اطلاعات هوشمند وجود ندارد باید به یکی از دو صورت «نظام هوشمند بازیابی اطلاعات» یا «نظام بازیابی هوشمند اطلاعات» تبدیل شود. همچنین عبارت «نظام بازیابی اطلاعات قراردادی» باید «نظام سنتی بازیابی اطلاعات» باشد زیرا مفهومی به‌نام «اطلاعات قراردادی» اصولاً نامفهوم است. عنوان فصل سوم نیز باید «نقش پردازش زبان طبیعی در نظام هوشمند بازیابی اطلاعات» باشد و این ترکیب باید در سرتاسر متن رعایت شود. این نکته در مورد «بازیابی اطلاعات قراردادی» نیز صادق است. در پاراگراف آخر ص. ۱۷ نیز عبارت «پردازش زبان طبیعی خودکار» باید «پردازش خودکار زبان طبیعی» باشد. در ص. اول ص. ۱۸، عبارت «اختلاف زبانی» باید «گونه‌گونی زبانشناختی» ترجمه شود؛ معادل انگلیسی این اصطلاح، Linguistic variation است. همین‌طور است ترجمه پانویس شماره ۱ در ص. ۲۴ که در آن، عبارت «(فراوانی مدرک، فراوانی مدرک معکوس)» باید «فراوانی اصطلاح، فراوانی معکوس مدرک» باشد زیرا معادل انگلیسی این اصطلاح عبارت است از: Term Frequency, Inverse Document Frequency. واژه Gerund در پانویس ص. ۴۸، بهتر است «اسم فعل» یا «اسم مصدر» ترجمه شود. در پانویس ۱ در ص. ۷۵، Bayes Theorem باید «اسم فعل» یا «اسم مصدر» ترجمه شود. در ص. ۴، س. ۱۵-۱۶ نمونه‌ای از ترجمه ناسلیس دیده می‌-

شود: "با وجود این، بخش تاثیر [تأثیر] ریشه سازی در بازیابی ناچیز نمی باشد." جمله "امتیازی که جداسازی مفاهیم مرکب دارد مثبت می باشد" در س. ۲ از ص. ۸ نیز از همین قبیل است. در ص. ۹ نیز با این جمله روبه‌رو می شویم: "وورهایس (۱۹۹۳) دریافت وقتی که از داده‌های انگلیسی با این روش استفاده می شود دقت ۳ امتیازی بین ۵ تا ۱۷٪ کاهش می یابد." جمله پس از این جمله هم اشکال دارد: "ولک و همکاران او (۲۰۰۲) آزمون مشابهی را بر روی داده‌های آلمانی و انگلیسی انجام داد[ند] و نتایج به هم آمیخته‌ای را گزارش نمود[ند]." منظور از «به هم آمیخته» در این جمله روشن نیست. در س. ۲ از ص. ۱۲ ترکیب «مزایای خوبی» دیده می شود: آیا مزایای بد هم وجود دارند؟! در س. ۱۳ از ص. ۳۴، می توان به جای استفاده از واژه نامناسب «با اثر»، واژه «مؤثر» را به کار برد. «انفجار پایگاه‌های اطلاعاتی غیرکتابشناختی شایع» هم از آن دسته جملاتی است که پیرامون آن باید بسیار اندیشید! در س. ۴ از ص. ۳۵، مفهوم این جمله در ص. ۳۱ روشن نیست: "ترویج بین نظام بازیابی اطلاعات و فنون هوش مصنوعی، طبیعی و به صورت بالقوه با اهمیت [بااهمیت] می باشد." منظور از واژه «توسعه‌های» در عبارت "توسعه‌های گوناگونی نسبت به فناوری نمایه معکوس استاندارد پیشنهاد شده است." در س. ۳ از ص. ۳۳ مبهم است. در همین صفحه، عبارت *Inverted index* بهتر است به جای «نمایه معکوس»، «نمایه مقلوب» ترجمه شود. در ص. ۳۶ نیز به این جملات بر می خوریم که درک آن‌ها دشوار است:

با وجود این، اسناد الزاماً به شکل زبان طبیعی یا متن آزاد که بازیابی اطلاعات هوشمند [بازیابی هوشمند اطلاعات] برای استنتاج و پی بردن به توان بازیابی اطلاعات نیاز دارد موجود و قابل استفاده می باشند.

در محیط بازیابی اطلاعات هوشمند [بازیابی هوشمند اطلاعات] مولفه‌ای [مولفه‌ای] زبانشناختی می تواند روابط اصطلاح را بین واحدهای نمایه سازی که ممکن است برای تولید توصیف های کامل تر مدرک نسبت به توصیف حاصل از اصطلاحات واحد به کار برد نشان دهد.

ترویج بین نظام بازیابی اطلاعات و فنون هوش مصنوعی که شامل پردازش زبان طبیعی، بازنمون دانش، استدلال و استنتاج می باشد طبیعی و بالقوه مفید و با اهمیت [بااهمیت] است.

در س. ۱۷-۱۸ از ص. ۴۲ نیز جمله زیر نشان از نبود سلاست در ترجمه دارد: "همچنین، از EAGLE برای نرمال سازی اشکال تکواژشناختی واژه‌ها استفاده کردیم، طوری

که واژه "foot" با "feet". تطبیق خواهد کرد. " مفهوم «تطبیق» در این جمله مبهم است. همچنین در س. ۵ از ص. ۴۳ جمله " برای رعایت کردن پرس‌وجوهای کوتاه آن گونه که در جستجوی وب به کار می‌روند، آزمون نظام خود را در محیط واقعی انجام دادیم. " نامفهوم است. همین گونه هستند سطرهای ۲-۴ از آخر در همین صفحه که عبارتند از: " از این طریق، برای هر پرس‌وجو تعداد ۱۸۰ سند به دست آوریم [آوردیم] که درباره نتایج حاصل از یک موتور کاوش وب نیز مصداق دارد. " در مورد شماره ۴ در ص. ۴۸، در جمله " ۴. ریشه‌های واژه را تنها با نادیده گرفتن مورفولوژی و برچسب‌های اجزای کلام تطبیق دهید. " معلوم نیست که ریشه‌ها را باید با چه تطبیق داد. ضمناً واژه مورفولوژی را می‌توان «ریخت‌شناسی» ترجمه کرد. این واژه «تطبیق» در موارد ۱ و ۵ در همین صفحه نیز ابهام برانگیز است. درک مفهوم این جمله در سطرهای ۳-۴ از ص. ۵۴ هم دشوار است: "مانند اسامی انگلیسی [اسامی انگلیسی]، اسامی فارسی [اسامی فارسی] به این جهت افزوده می‌شوند تا مالکیت، تعدد و چند تایی [چندتایی] را معنی بخشند. " در س. ۴-۵ کتاب نیز، جمله‌ای ناسلیس دیده می‌شود: «با وجود این، به ندرت اتفاق می‌افتد که ریشه‌ساز بیش از حد واژه را حذف کند و به ترکیب یا آمیختگی دو واژه که دو جور به نظر رسیده است باعث شود.» در سطر دوم از ص. ۸۵، جمله «پیشاپیش، واژه‌های تهی برطرف شده بودند.» از نوع جملات بدون سلاست است.

اشکالاتی نیز در کتاب وجود دارند که درک برخی از مفاهیم مطرح در آن را دشوار می‌سازند. برای نمونه در فصل دوم، مثال ۵ زمانی کارایی دارد که پاراگراف مطرح در آن، به انگلیسی آورده شده باشد. همچنین معلوم نیست چرا جدول‌ها و شکل‌های فصل سوم ترجمه نشده‌اند؛ این‌ها دارای متن‌های ساده‌ای هستند و به درک بهتر مطلب کمک می‌کنند. در جدول ۱ در فصل سوم، اولاً واژه test باید text باشد و ثانیاً در خانه چهارم زیر IIR افعال باید احتمالاً به‌حالت سوم شخص مفرد باشند. اشتباهات این جدول جای بحث فراوانی دارد. در جدول ۲ در ص. ۵۵، در حالی فعل «کردن» به‌عنوان یک فعل بی‌قاعده آمده است که در متن پیش از آن، قید شده که "مصدرهای با قاعده [باقاعده] در زبان فارسی به پسوند "دن" ختم می‌شود؛ ضمن این که مثال «کردن» در جدول ۲، ناقص ادعایی است که در متن پیش از جدول ۱ آمده است: "مصدرهای با قاعده [باقاعده] در زبان فارسی به پسوند "دن" ختم می‌شود و وجه امری مصدرهای با قاعده [باقاعده] با حذف دو یا گاهی سه کاراکتر آخر به [آخر به] دست می‌آید." در همین جا می‌توان به فعل «دادن» نیز

اشاره کرد که وجه امری آن «ده» و ناقض ادعای مطرح در این صفحات است. در نوشته زیر جدول ۶ در ص. ۵۹، به ردیف "s" و ستون "1" اشاره شده اما این حرف و عدد در هیچ یک از تصاویر دیده نمی‌شوند. در س. ۱۴-۱۷ ص. ۷۹ این جمله دیده می‌شود: "بنابراین، قرار دادن سیستمی از یک سازو کار [سازوکار یا ساز و کار] که استاندارد کردن واژه‌های مختلف را امکان پذیر سازد و بدین ترتیب اشکال مختلف یک ریشه واحد را زیر یک شکل واحد که ممکن است در پرس‌وجو و در مدرک وجود داشته باشند [باشد] ضروری می‌باشد." این جمله هم نامفهوم و هم ناسلیس می‌باشد. در ص. ۳۴ کتاب نیز در انتهای پاراگراف اول ادعایی مطرح شده که عجیب می‌نماید: «چون پایگاه‌های اطلاعاتی غیرکتابشناختی از تعداد آثار کتابشناختی چاپی به مراتب بیشتر شده است، نیاز به بازیابی اطلاعات خودکار [بازیابی خودکار اطلاعات] یا بازیابی اطلاعات هوشمند [بازیابی هوشمند اطلاعات] افزایش یافته است.» آیا واقعاً پایگاه‌های اطلاعاتی غیرکتابشناختی از تعداد آثار کتابشناختی چاپی به مراتب بیشتر شده‌اند؟ در ص. ۴۶ از مفهوم «پنج‌ره سه جمله‌ای» استفاده شده که نامفهوم و قبلاً درباره آن توضیحی داده نشده است. در س. ۱۸-۱۹ از ص. ۹۴، این پرسش مطرح می‌شود که آیا ریشه‌سازی واژه انگلیسی computing، ریشه compact را به وجود می‌آورد.

یکی از ایرادهای دیگر وارد به کتاب، نبود پانویس برای نام‌های خارجی است: مثلاً برای نام‌های هارمان و کروتر در ص. ۴ پانویس داده نشده است. در هیچ موردی برای نام‌های خارجی پانویس دیده نمی‌شود.

فهرست‌های منابع نیز بی‌ایراد نیستند. مثلاً شناسه مربوط به Harman در ص. ۱۴ دارای اشکال و شکلی غیرعادی است. ضمن این که در مواردی به نامی در متن استناد شده که در فهرست منابع آن فصل وجود ندارد: هال و همکاران (۱۹۹۶) در ص. ۸ نمونه‌ای از این موارد است. در مواردی استناد درون متنی با شناسه مربوط در ارجاعات همخوانی ندارد: (Manning, 1999) در ص. ۲۱ که معلوم نیست چرا به انگلیسی نوشته شده در ارجاعات، دارای نویسنده همکار است که در استناد درون متنی ذکر نشده است. در ارجاعات فصل دوم، شناسه‌های مربوط به Baeza-Yates به شکلی غیرعادی هستند. در ارجاعات برخی فصل‌ها، میان صفحات یک مقاله، به جای خط تیره، از حرف t استفاده شده است. این اشکال در بیشتر ارجاعات فصل دو دیده می‌شود. منبع شماره [۲] در ص. ۹۹، ظاهراً دو منبع است که تجانسی با هم ندارند. این اشکالات باید بازنگری و اصلاح شوند.

برخی اصلاحات دیگر هم در کتاب لازم است: مثلاً جای جدول ۳ باید پس از شماره ۴ در ص. ۱۰ باشد. در شکل ۱ در ص. ۱۲، ظاهراً IE باید IR باشد. در همین شکل، عبارت «بازیابی اطلاعات» استفاده شده اما در ص. ۱۳، در متن مربوط به همین شکل، عبارت «بازیابی نقل قول» به کار رفته که درست نیست. این دو عبارت معادل Passage retrieval گرفته شده‌اند که هیچ کدام درست نیستند.

۴. برخی ایرادهای جزئی

ایرادهای جزئی هم در کتاب دیده می‌شوند: در ص. ۴ دو پانویس وجود دارند که دارای شماره‌های یکسانی می‌باشند. اشتباهات تایپی هم در زمره این ایرادهای جزئی هستند: در شکل ۱ در فصل دوم، واژه Cunction باید Function باشد. در ص. ۴۵، فرمولی برای Relevance (ربط) وجود دارد که در آن، ظاهراً واژه token (به معنی نماد) چهار بار به یکی از صورت‌های زیر اشتباه نوشته شده است: taken، tokanes و «پرس و جو» هم به شکل‌های مختلفی نوشته شده است: «پرس و جو»، «پرس و جو»، «پرس و جو»، و «پرس و جو». این‌ها باید یکدست شوند. بهتر است میان دو واژه مترادف «نظام» و «سیستم»، یکی انتخاب و در سرتاسر کتاب استفاده شود. برخی از ارقام درون قلاب در فصل چهارم (مانند آن چه که در ص. ۶۰ آمده، با شماره‌های مربوط به ارجاعات اشتباه می‌شوند. تلفظ برخی نام‌های خارجی نیز اشتباه است: نام «بوکشتاین» در ص. ۷۵ باید «بوکستاین» باشد. در واژه‌نامه که از نکات مثبت کتاب است ۳۳ اشتباه وجود دارد. برخی ناهماهنگی‌ها در ترجمه نیز در واژه‌نامه مشخص شده‌اند: در واژه‌نامه، Dictionary-based «فرهنگ محور» ترجمه شده که درست هم نیست؛ در عین حال، Memory-based «حافظه بنیاد» ترجمه شده؛ همین‌طور Rule-based که «قاعده بنیاد» است.

۵. رابطه میان اصطلاحات ترجمه شده در متن، واژه‌نامه و نمایه

رابطه میان اصطلاحات ترجمه شده در متن، واژه‌نامه و نمایه در این کتاب به گونه‌ای است که بررسی مفصلی را می‌طلبد. به همین دلیل بخشی از نقد حاضر به این رابطه اختصاص داده شده است. این بخش از این نظر که می‌تواند بار آموزشی نیز برای نمایه‌سازان و مترجمان داشته باشد در خور توجه است. این رابطه از طریق دو جدول زیر نمایش داده

می‌شود. جدول ۱ رابطه میان اصطلاحات ترجمه شده در متن، واژه‌نامه و نمایه را نشان می‌دهد.

جدول ۱: رابطه میان اصطلاحات ترجمه شده در متن، واژه‌نامه و نمایه

ردیف	اصطلاح در متن	معادل انگلیسی در متن	معادل انگلیسی در واژه‌نامه	برابری معادل‌های انگلیسی در متن و واژه‌نامه	معادل فارسی در واژه‌نامه	برابری معادل‌های فارسی در متن و واژه‌نامه	اصطلاح معادل در نمایه	تعداد واژگان در واژه‌نامه و نمایه	برابری تعداد واژگان
۱	پردازش زبان طبیعی	Natural Language Processing (NLP)	Natural Language Processing	✓	پردازش زبان طبیعی	✓	-	-	✓
۲	بازیابی اطلاعات	Information Retrieval (IR)	-	-	-	-	-	-	-
۳	تعیین مقوله دستوری	Part-of-Speech Tagging	-	-	-	-	-	-	-
۴	دقت بازیافت با ۱۱ امتیاز؛ دقت با میانگین ۱۱ امتیاز	11 pt [pt.] Precision	11 pt [pt.] Precision	✓ on here.	دقت بازیافت با ۱۱ امتیاز	✓	دقت بازیافت با ۱۱ امتیاز	-	-
۵	دقت بازیافت R	R-Precision	R-Precision	✓	دقت بازیافت R	✓	دقت بازیافت R	✓	✓
۶	مدل فضای برداری	Vector Space Model	Vector Space Model	✓	مدل فضای برداری	✓	مدل فضای برداری	✓	✓
۷	ریشه‌سازی تصریفی	Inflectional Stemming	-	-	-	-	ریشه‌سازی تصریفی	-	-
۸	ریشه‌سازی اشتقاقی	Derivational Stemming	Derivational Stemming	✓	ریشه‌سازی اشتقاقی	✓	ریشه‌سازی اشتقاقی	-	-
۹	مقوله نحوی؛ مقوله دستوری	Syntactic Category	Syntactic Category	✓	مقوله نحوی	✓	مقوله نحوی؛ مقوله دستوری*	-	-
۱۰	حافظه-بنیاد	Memory-based	Memory-based	✓	حافظه بنیاد	✓	-	-	-
۱۱	قاعده-بنیاد	Rule-based	Rule-based	✓	قاعده بنیاد	✓	-	-	-
۱۲	اسمارت	SMART	SMART	✓	اسمارت (زبان طبیعی اسمارت)	✓	اسمارت	✓	✓

۱۳	تقطیع واژه‌ها	Word Segmentation	Word Segmentation	تقطیع واژه	تقطیع واژه‌ها	✓
۱۴	بایگرام‌ها	Bigram	Bigrams	بایگرام	-	
۱۵	فرهنگ-محور	Dictionary-based	Dictionary-based	فرهنگ محور	-	
۱۷	فرایند باز آور	Productive Process	Productive Process	فرایند باز آور	-	✓
18	واژگان	Lexicon	Lexicon	واژگان	✓	
۱۹	تشریح (تجزیه) دستوری سطحی	Shallow Parsing	Shallow Parsing	تجزیه دستوری سطحی	-	✓
۲۰	ماشین‌های بردار پشتیبان	Support Vector Machine	Support Vector Machines	ماشین بردار پشتیبان	-	
۲۱	جفت‌های سر- تعدیل‌گر	Head Modifier Pairs [Pairs]	Head-Modifier pairs	جفت‌های سر تعدیل‌گر	جفت‌های سر- تعدیل‌گر	
۲۲	تجزیه‌گر تابع	Dependency Parser	Dependency Parser	تجزیه‌گر تابع	-	✓
۲۳	ابهام زدایی از معنی واژه‌ها	Word-Sense Disambiguation	Word-sense Disambiguation	ابهام زدایی از مفهوم واژه	ابهام زدایی از واژه‌ها	✓
۲۴	تجزیه‌گر حالت نامتناهی*	Finite State Parser	Finite State Parser	تجزیه‌گر حالت نامتناهی	-	✓
۲۵	مفعول صریح	Direct Object	Direct Object	مفعول صریح	-	✓
۲۶	صفت فرعی	Adjunct	Adjunct	صفت فرعی	-	✓
۲۷	بازیابی مدرک	Document Retrieval	Document Retrieval	بازیابی مدرک	-	✓
۲۸	بازیابی عبارت	Passage Retrieval	Passage Retrieval	بازیابی عبارت	-	✓
۲۹	پرسش و پاسخ	Question Answering	Question Answering	پرسش و پاسخ	-	✓
۳۰	سفارشی	Tailor Mode [Made]	Tailor Made	سفارشی	-	✓
۳۱	چند معنایی	Polysemy	Polysemy	چند معنایی	چند معنایی	✓
۳۲	هم معنایی	Synonymy	Synonymy	هم معنایی	هم معنایی	✓
۳۳	اختلاف زبانی	Linguistic Variation	Linguistic Variation	اختلاف زبانی	✓	
۳۴	اعوجاج مدرک	Document Noise	Document Noise	اعوجاج مدرک	اعوجاج مدرک	✓
۳۵	سطح تکواژ شناختی	Morphological Level	Morphological Level	سطح تکواژ شناختی	سطح تکواژ شناختی	✓
۳۶	سطح نحوی	Syntactic Level	Syntactic Level	سطح نحوی	سطح نحوی	✓
۳۷	اختلاف واژگان	Lexicon Variation	Lexicon Variation	اختلاف واژگان	اختلاف واژگان	✓

۳۸	سطح منظور شناختی	Pragmatic Level	√	سطح منظور شناختی	√	سطح منظور شناختی	√
۳۹	تفسیر لفظی	Literal Interpretation	√	تفسیر لفظی	√	تفسیر لفظی	√
۴۰	ربط	Relevance	√	ربط	√	ربط	√
۴۱	نمایه اصطلاحات	Terms Index	√	نمایه اصطلاحات	√	-	-
۴۲	بسته‌ای از واژه‌ها	Bag of Words	√	بسته‌ای از واژه‌ها	√	-	-
۴۳	مولفه‌ها	Parameters	√	مولفه	√	-	-
۴۴	اصطلاحات مربوط	Relevant Terms	√	اصطلاحات مرتبط	√	-	-
۴۵	سرآیندها	Headers	-	سرآیند	-	-	-
۴۶	سرنام‌ها	Acronyms	√	سرنام	√	-	-
۴۷	حروف اضافه	Prepositions	-	حروف اضافه	√	-	-
۴۸	شکل‌های تصریفی	Inflected Forms	-	شکل‌های تصریفی	-	-	-
۴۹	فراوانی مدرک، فراوانی مدرک معکوس	Term Frequency, Inverse Document Frequency	-	فراوانی مدرک	-	-	-
۵۰	ارزش تمیز دهنده	Discriminatory Value	√	ارزش تمیز دهنده	√	-	-
۵۱	فهرست واژه-های غیرمجاز	Stop words [Stop words] List	-	واژه غیر مجاز	-	واژه‌های غیرمجاز	-
۵۲	تجزیه کننده‌ها	Parsers	-	تجزیه کننده	-	تجزیه کننده‌ها	-
۵۳	جمله‌های اسمی	Nominal Sentences	√	جملات اسمی	√	-	-
۵۴	جملات اضافی	Prepositional Sentences	√	جملات اضافی	√	-	-
۵۵	پایگاه اطلاعاتی واژگانی وردنت	Lexicographic Database WordNet	√	پایگاه اطلاعاتی واژگانی وردنت	√	وردنت	-
۵۶	مجموعه‌های مترادف	Synsets	√	مجموعه‌های مترادف	√	-	-
۵۷	موجودیت‌ها	Entities	√	موجودیت	√	موجودیت‌ها	-
۵۸	پایگاه اطلاعاتی هستی‌شناسی	Ontology Database	√	هستی‌شناسی	√	پایگاه اطلاعاتی هستی‌شناسی	-
۵۹	نظام زبان طبیعی	SMART	√	اسمارت (زبان طبیعی)	√	اسمارت	-

نسبت‌سنجی تریلوژی آیسخولوس و آپولوژی سقراط از ره‌گذر ... (سیدمجید کمالی و دیگران) ۴۰۱

			اسمارت)				اسمارت
√	میان زبانی	√	میان زبانی	√	Interlingual	Interlingual	۶۰ میان زبانی
	-		-		-	Empty Phrases	۶۱ عبارات‌های تهی
	-	√	قراردادی	√	Conventional	Conventional	۶۲ قراردادی
√	هوش مصنوعی	√	هوش مصنوعی	√	Artificial Intelligence	Artificial Intelligence	۶۳ هوش مصنوعی
√		√	فوق متن		Hyper Text	Hypertext	۶۴ فوق متن
√	بردارهای اصطلاح		بردار اصطلاح		Term Vector	Term Vectors	۶۵ بردارهای اصطلاح
√	اپراتورهای بولی	√	اپراتورهای بولی		Boolean Operator	Boolean Operators	۶۶ اپراتورهای بولی
√	نمایه معکوس	√	نمایه معکوس	√	Inverted Index	Inverted Index	۶۷ نمایه معکوس
	-	√	زبان طبیعی	√	Natural Language	Natural Language	۶۸ زبان طبیعی
√	متن آزاد	√	متن آزاد		Free Text	Free-text	۶۹ متن آزاد
√	واژه‌های هم‌نویسه	√	واژه‌های هم‌نویسه	√	Homographs	Homographs	۷۰ واژه‌های هم‌نویسه
	-		تفسیر منظورشناختی	√	Pragmatic Interpretation	Pragmatic Interpretation	۷۱ تفسیر منظورشناختی
	شناسگرها		شناسگر	√	Recognizer	Recognizer	۷۲ شناسگر
	رتبه‌بندها		رتبه‌بند	√	Ranker	Ranker	۷۳ رتبه‌بند
	-		ماجول	√	Module	Module	۷۴ ماجول‌ها
	-	√	رشته	√	String	String	۷۵ رشته
	-	√	عبارت اسمی		Noun-Phrase	Noun-phrase	۷۶ عبارت اسمی
√	شمول معنایی	√	شمول معنایی	√	Hyponymy	Hyponymy	۷۷ شمول معنایی
	-		بسته وردنت پرینستون	√	Princeton Word Net Package	Princeton Word Net Package	۷۸ بسته وردنت پرینستون [پرینستون]
	-		زیر مفهوم		Sub Concept	Sub-concept	۷۹ زیر مفهوم
√	آلتاویستا	√	آلتا ویستا	√	Alta Vista	Alta Vista	۸۰ آلتا ویستا
	-	√	فضای نمادی	√	Token Space	Token Space	۸۱ فضای نمادی
	-		نسخه واینای فضای برداری	√	Vanila Version of Vector Space	Vanila Version of Vector Space	۸۲ نسخه " واینای" ی فضای برداری
√	همجواری	√	همجواری	√	Proximity	Proximity	۸۳ همجواری
√	جفت جلویی	√	جفت جلویی		Forward Paris [Pairs]	Forward Pairs	۸۴ جفت جلویی
√	جفت عقبی	√	جفت عقبی		Backward Paris [Pairs]	Backward Pairs	۸۵ جفت عقبی
	-		ثبت		Register	Registers	۸۶ ثبت‌ها
	-	√	پشتیبانی		Backing off	Backing Off	۸۷ پشتیبانی
	-		مصدر فعلی		Gerund	Gerunds	۸۸ مصدرهای

۴۰۲ پژوهش‌نامه انتقادی متون و برنامه‌های علوم انسانی، سال بیستم، شماره دهم، دی ۱۳۹۹

							فعلی	
	شناسگرها		شناسگر عبارت اسمی	√	Noun Phrase Recognizer	Noun Phrase Recognizer	شناسگر عبارت اسمی	۸۹
	-		-		-	Farsi	فارسی	۹۰
	-		ریشه ساز لاوین		Lovin Stemmer	Lovins Stemmer	ریشه ساز لاوینز	91
	√	ریشه ساز پورتر	ریشه ساز پورتر		Porter Stemmer	Porter Stemme [Stemmer]	ریشه ساز پورتر	۹۲
	√	دقت	دقت	√	Precision	Precision	دقت	۹۳
	√	بازیافت	بازیافت	√	Recall	Recall	بازیافت	۹۴
	-		پسوندها		Suffix	Suffixes	پسوندها	۹۵
	-		پیشوندها		Prefix	Prefixes	پیشوندها	۹۶
	-	√	نفی	√	Negation	Negation	نفی	۹۷
	-	√	وجه	√	Mood	Mood	وجه	۹۸
	-	√	مصدر	√	Infinitive	Infinitive	مصدر	۹۰
	-	√	مضارع اختیاری	√	Indicative Present Tense	Indicative Present Tense	مضارع اختیاری	۹۱
	-	√	تفضیلی	√	Comparative	Comparative	تفضیلی	۹۲
	-		مصمت		Consonant	Consonants	مصمت‌ها	۹۳
	-		مصوت		Vowel	Vowels	مصوت‌ها	۹۴
	-	√	زیر رشته انتهاپی	√	Terminal Substring	Terminal Substring	زیر رشته انتهاپی	۹۵
	-	√	حالت اضافه جمع	√	Plural Possessive	Plural Possessive	حالت اضافه جمع	۹۶
	-		پسوندهای فعلی		Verb-Suffix	Verb- suffixes	پسوندهای فعلی	۹۷
	-		پسوندهای اسم جمع		Plural-Noun- Suffix	Plural-noun- suffixes	پسوندهای اسم جمع	۹۸
	-		پسوندهای اسم ملکی		Possessive- Noun-Suffix	Possessive- noun-suffixes	پسوندهای اسم ملکی	۹۹
	ماشین خودکار قطععی با حالت نامتناهی	√	ماشین خودکار قطععی با حالت متناهی		Deterministic Finite Automata	Deterministic Finite Automata (DFA)	ماشین خودکار قطععی با حالت متناهی	۱۰۰
	پس پردازشگر		پس پردازشگر	√	Post Processor	Post Processor	پس پردازشگر	۱۰۱
	-		بردار محور	√	Vector-based	Vector-based	بردار - محور	۱۰۲
	یورو فارسی		یورو فارسی	√	Euro Farsi	Euro Farsi	یوروفارسی	۱۰۳
	-	√	قطعه‌ای کوچک از کد که بر روی اینترنت قابل	√	Applet	Applet	قطعه‌ای کوچک از کد که بر روی اینترنت قابل	۱۰۴

نسبت‌سنجی تریلوژی آیسخولوس و آپولوژی سقراط از ره‌گذر ... (سیدمجید کمالی و دیگران) ۴۰۳

			انتقال باشد و بر روی ماشین دریافت کننده اجرا شود.				انتقال باشد و بر روی ماشین دریافت کننده اجرا شود.	
	-		مجموعه‌ای متوالی از بیت‌هایی که تصویری را در حافظه معرفی می‌کند که باید بر روی صفحه نمایش داده شود، به ویژه در سیستم‌هایی که رابط گرافیکی کاربر دارند.	√	Bitmap	Bitmap	مجموعه‌ای متوالی از بیت‌هایی که تصویری را در حافظه معرفی می‌کند که باید بر روی صفحه نمایش داده شود، به ویژه در سیستم‌هایی که رابط گرافیکی کاربر دارند.	۱۰۵
	-		فضای خالی: نواحی خالی روی یک صفحه که می‌توان آن را در یک طرح برای ایجاد توازن، وضوح و جلب نظر به کار برد.	√	White Space	White Space	نواحی خالی روی یک صفحه که می‌توان آن را در یک طرح برای ایجاد توازن، وضوح و جلب نظر به کار برد.	۱۰۶
	-	√	آغازی	√	Initial	Initial	آغازی	۱۰۷
	-	√	میانی	√	Medial	Medial	میانی	۱۰۸
	-	√	کاراکتر فاصله	√	Space Character	Space Character	کاراکتر فاصله	۱۰۹
	-		زیر مجموعه‌ای از زبان جاوا که در مقایسه با جاوا عملکرد ضعیفی دارد. برنامه‌های کاربردی فوری، اصلی و توابع را می‌توان با سند اصلی جاوا به	√	Java Script	Java Script	زیر مجموعه‌ای از زبان جاوا که در مقایسه با جاوا عملکرد ضعیفی دارد. برنامه‌های کاربردی فوری، اصلی و توابع را می‌توان با سند اصلی جاوا به	۱۱۰

			صفحات وب اضافه کرد.				صفحات وب اضافه کرد.	
			فرایندی که به فضای تعویض برنامه امکان می‌دهد. تا در صورت نیاز اضافه شود. این فرایند به یک فایل تعویضی اجازه می‌دهد تا در هر لحظه روی هر دیسک آماده ای ایجاد شود و فقط تا زمان خاموش شدن سیستم فعال باشد.	√	Generate on the Fly	Generate on the fly	فرایندی که به فضای تعویض برنامه امکان می‌دهد تا در صورت نیاز اضافه شود. این فرایند به یک فایل تعویضی اجازه می‌دهد تا در هر لحظه روی هر دیسک آماده‌ای ایجاد شود و فقط تا زمان خاموش شدن سیستم فعال باشد.	۱۱۱
	-		-	-	-	Unload	از حافظه برداشتن	۱۱۲
			سرویس دهنده وب HTTP تعبیه شده	√	Embedded HTTP Web Server	Embedded HTTP Web Server	سرویس دهنده وب HTTP تعبیه شده	۱۱۳
			سیستم تولید حروف	√	Type Foundry	Type Foundry	سیستم تولید حروف	۱۱۴
			گراف‌های شبکه قابل حمل	√	Portable Network Graphics	Portable Network Graphics	گراف‌های شبکه قابل حمل	۱۱۵
	√	ماشین خودکار با حالت متناهی	ماشین با حالت متناهی	√	Finite State Machine	Finite State Machine	ماشین خودکار با حالت متناهی	۱۱۶
			قواعد پشته بندی	√	Stacking	Stacking	قواعد پشته بندی	۱۱۷
			حاصلضرب داخلی	√	Inner Product	Inner Product	حاصلضرب داخلی	۱۱۸
			قضیه بیزی	√	Bayes Theorem [Theorem]	Bayes Theorem [Theorem]	قضیه بیزی	۱۱۹
			مدل سازی زبان آماری	√	Statistical Language Modelling	Statistical Language Modeling	مدل سازی زبان آماری	۱۲۰
			توزیع پواسون:	√	Poisson Distribution	Poisson Distribution	توزیع پواسون متحنی ریاضی	۱۲۱

نسبت‌سنجی تریلوژی آیسخولوس و آپولوژی سقراط از ره‌گذر ... (سیدمجید کمالی و دیگران) ۴۰۵

			منحنی ریاضی که غالباً در آمار و شبیه سازی از آن استفاده می شود تا صحت وقوع رویدادها را نشان دهد.				که غالباً در آمار و شبیه سازی از آن استفاده می شود تا صحت وقوع رویدادها را نشان دهد.	
			فنون نرم روان	√	Smoothing Techniques	Smoothing Techniques	فنون نرم و روان	۱۲۲
	-	√	الگوریتم داوسون	√	Dawson	Dawson	الگوریتم داوسون	۱۲۳
	-		ریشه‌ساز پاپس		Paice Stemmer	Paice	الگوریتم پاپس	۱۲۴
	-		واج گونه	√	Allomorph	Allomorph	واج گونه	۱۲۵
	-	√	حروف میانه	√	Intermediate Letters	Intermediate Letters	حروف میانه	۱۲۶
	-		مسند فعلی	√	Verbal Predicates	Verbal Predicates	مسندهای فعلی	۱۲۷
	-	√	مصوت بی تکیه		Unstressed vowel	Unstressed Vowel	مصوت بی تکیه	۱۲۸
			غیر-اسکی		Non-ASCII	Non - ASCII	غیر-اسکی	۱۲۹
			سرورس دهنده پردازش		Server Processing	Server - Processing	سرورس دهنده - پردازش	۱۳۰
			علامت اختصاری (استانداردی که امکان انتقال داده-هایی مانند تصویر، صدا، فایل های دودویی را توسط پست الکترونیکی شبکه اینترنت بدون هیچ اجباری برای تبدیل آن با فرمت ASCII فراهم می-سازد)		MIME (Multipurpose Internet Mail Extensions)	MIME	MIME علامت اختصاری Multipurpose Internet Mail Extensions استاندارد که امکان انتقال داده های مانند تصویر، صدا و فایل های دودویی را توسط پست الکترونیکی شبکه اینترنت بدون هیچ اجباری برای تبدیل آن با فرمت ASCII	۱۳۱

							فراهم می‌سازد.	
۱۳۲	کوتاه سازی	Truncation	Truncation	√	کوتاه سازی	√	کوتاه سازی	√
۱۳۳	ترتیب واژه با شکل آزاد	Free Word Order	Free word-order	√	ترتیب واژه با شکل آزاد	-		
۱۳۴	اشتقاقی	Derivative	Derivative	√	اشتقاقی	√		
۱۳۵	واژه جامد	Primitive	Primitive	√	واژه جامد	√		
۱۳۶	اسامی فاعلی و مفعولی	-	Participles	-	-	-		
137	ضمایر موصولی	Reletives [Relatives]	Relatives	√	ضمایر موصولی	√		
۱۳۸	ضمیمه اشاره	Demonstrative	Demonstrative	√	ضمیمه اشاره	√		
۱۳۹	ادات پرسش	Interrogative	Interrogatives	√	ادات پرسش	√		
۱۴۰	کامل	Perfect	Perfect	√	کامل	√		
۱۴۱	استمراری	Imperfect	Imperfect	√	استمراری	√		
۱۴۲	وجه امری	Imperative	Imperative	√	وجه امری	√		
۱۴۳	ادات فعل	Particle	Particle	√	ادات فعل	√		
۱۴۴	واژه‌های تعجبی (Interjection) مانند: n: واژه‌هایی مانند ugh! ، gosh! یا Wow که بیانگر نوعی حالت عاطفی یا نگرشی می‌باشد	Interjection	Interjection	√	واژه‌های تعجبی مانند: Wow! ، gush! یا uyh! که بیانگر نوعی حالت عاطفی یا نگرشی می‌باشد	√		
145	پایه کلمه	Root	Root	√	پایه کلمه	√		
۱۴۶	ریشه کلمه	Stem	Stem	√	ریشه کلمه	√		
۱۴۷	مالکیت مدرک	Document Privacy	Document Privacy	√	مالکیت مدرک	√		
۱۴۸	آین	Ayna	Ayna	√	آین	√		
جمع							۱۴۴	۱۴۸

*در نمایه «مقوله دستوری» وجود دارد.

** این عبارت باید «تجزیه‌گر حالت متناهی» باشد. کلاً برخی از معادل‌ها اشتباه به نظر

می‌رسند.

در جدول بالا، واژگان اشتباه در قلاب اصلاح شده‌اند.

نابرابرها در جدول بالا (آنهایی که دارای نشانه √ نیستند) شامل: نبود معادل، تفاوت در

فاصله‌های پیش و پس از حروف و نشانه‌ها، تفاوت در جمع و مفرد بودن واژگان و یا

نسبت‌سنجی تریلوژی آیسخولوس و آپولوژی سقراط از ره‌گذر ... (سیدمجید کمالی و دیگران) ۴۰۷

عبارت، بود یا نبود خطّ تیره میان دو واژه یا عبارت، اشتباهات تایپی که به‌نوعی باعث تفاوت شده‌اند، بزرگی و کوچکی حروف آغازین در واژگان یا عبارت‌های انگلیسی، چسبیده بودن یا نبودن برخی از واژگان، ارائه تعداد متفاوت از معادل‌ها در جاهای مختلف، یا مجموعه‌ای از دو یا چند تا از این‌هاست.

برای ۱۴۸ واژه در متن، معادل اصلی داده شده است. اما چهار واژه از این معادل‌ها در واژه‌نامه دیده نمی‌شوند و برای این کار نمی‌توان دلیل مشخصی یافت. ۱۰۵ واژه یا عبارت انگلیسی (حدود ۷۰/۹۴٪) در واژه‌نامه عیناً با معادل انگلیسی مربوط در متن، مشابه هستند. معادل‌های فارسی ۸۶ واژه یا عبارت (حدود ۵۸/۱٪) در واژه‌نامه عیناً با معادل فارسی مربوط در متن همخوانی دارند. تنها ۷۶ واژه یا عبارت در نمایه وجود دارند و این‌ها تنها ۵۲/۷٪ از واژگان و عبارات واژه‌نامه را تشکیل می‌دهند.

جدول ۲ رابطه میان اصطلاحات ترجمه شده در متن، واژه‌نامه و نمایه بدون توجه به معادل‌های انگلیسی در متن، واژه‌نامه، و نیز، برابری معادل‌های انگلیسی در متن و واژه‌نامه را نشان می‌دهد.

جدول ۲: رابطه میان اصطلاحات ترجمه شده در متن، واژه‌نامه و نمایه بدون توجه به معادل‌های انگلیسی در متن، واژه‌نامه، و برابری معادل‌های انگلیسی در متن و واژه‌نامه

ردیف	اصطلاح در متن	معادل فارسی در واژه‌نامه	اصطلاح معادل در نمایه	برابری شکل واژگان و عبارات نمایه با واژه‌نامه	برابری شکل واژگان و عبارات نمایه با متن
1	دقت بازیافت با ۱۱ امتیاز؛ دقت با میانگین ۱۱ امتیاز	دقت بازیافت با ۱۱	دقت بازیافت با ۱۱ امتیاز	√	
2	دقت بازیافت R	دقت بازیافت R	دقت بازیافت R	√	√
3	مدل فضای برداری	مدل فضای برداری	مدل فضای برداری	√	√
4	ریشه سازی تصریفی	-	ریشه‌سازی تصریفی		
5	ریشه سازی اشتقاقی	ریشه سازی اشتقاقی	ریشه‌سازی اشتقاقی		
6	مقوله نحوی؛ مقوله دستوری	مقوله نحوی	مقوله نحوی نک مقوله های دستوری*		
7	اسمارت	اسمارت (زبان طبیعی اسمارت)	اسمارت		√
8	تقطیع واژه‌ها	تقطیع واژه	تقطیع واژه‌ها		√
9	جفت‌های سر- تعدیل‌گر	جفت‌های سر تعدیل‌گر	جفت‌های سر-تعدیل‌گر		

10	ابهام زدایی از مفهوم واژه، ابهام زدایی از معانی واژه- ها	ابهام زدایی از مفهوم واژه	ابهام زدایی از واژه‌ها		
11	چند معنایی	چند معنایی	چند معنایی	✓	✓
12	هم معنایی	هم معنایی	هم معنایی	✓	✓
13	اعوجاج مدرک	اعوجاج مدرک	اعوجاج مدرک	✓	✓
14	سطح تکواژ شناختی	سطح تکواژ شناختی	سطح تکواژ شناختی		✓
15	سطح نحوی	سطح نحوی	سطح نحوی	✓	✓
16	اختلاف واژگان	اختلاف واژگان	اختلاف واژگان	✓	✓
17	سطح منظور شناختی	سطح منظور شناختی	سطح منظور شناختی		✓
18	تفسیر لفظی	تفسیر لفظی	تفسیر لفظی	✓	✓
19	ربط	ربط	ربط	✓	✓
20	فهرست واژه‌های غیرمجاز	واژه غیر مجاز	واژه‌های غیرمجاز		
21	تجزیه کننده‌ها	تجزیه کننده	تجزیه کننده‌ها		
22	پایگاه اطلاعاتی واژگانی وُردنیت	پایگاه اطلاعات واژگانی وردنت	وردنت		
23	موجودیت‌ها	موجودیت	موجودیت‌ها		
24	پایگاه اطلاعاتی هستی- شناسی	هستی شناسی	پایگاه اطلاعاتی هستی شناسی		
25	نظام زبان طبیعی اسمارت	اسمارت (زبان طبیعی اسمارت)	اسمارت		
26	میان زبانی	میان زبانی	میان زبانی	✓	✓
27	هوش مصنوعی	هوش مصنوعی	هوش مصنوعی	✓	✓
28	فوق متن	فوق متن	فوق متن	✓	✓
29	بردارهای اصطلاح	بردار اصطلاح	بردارهای اصطلاح		✓
30	اپراتورهای بولی	اپراتورهای بولی	اپراتورهای بولی	✓	✓
31	نمایه معکوس	نمایه معکوس	نمایه معکوس	✓	✓
32	متن آزاد	متن آزاد	متن آزاد	✓	✓
33	واژه‌های هم نویسه	واژه‌های هم نویسه	واژه‌های هم نویسه	✓	✓
34	شناسگر	شناساگر	شناسگرها		
35	رتبه‌بند	رتبه بند	رتبه بندها		
36	شمول معنایی	شمول معنایی	شمول معنایی	✓	✓
37	آلتا ویستا	آلتا ویستا	آلتا ویستا		✓
38	همجواری	همجواری	همجواری	✓	✓
39	جفت جلویی	جفت جلویی	جفت جلویی	✓	✓
40	جفت عقبی	جفت عقبی	جفت عقبی	✓	✓
41	شناسگر عبارت اسمی	شناساگر عبارت اسمی	شناسگرها		

نسبت‌سنجی تریلوژی آیسخولوس و آپولوژی سقراط از ره‌گذر ... (سیدمجید کمالی و دیگران) ۴۰۹

42	ریشه‌ساز پورتر	ریشه‌ساز پورتر	ریشه‌ساز پورتر	√	
43	دقت	دقت	دقت	√	√
44	بازیافت	بازیافت	بازیافت	√	√
45	ماشین خودکار قطعی با حالت متناهی	ماشین خودکار قطعی با حالت متناهی	ماشین خودکار قطعی با حالت نامتناهی		
46	پس-پردازشگر	پس پردازشگر	پس-پردازشگر		
47	یوروفارسی	یورو فارسی	یورو فارسی	√	
48	ماشین خودکار با حالت متناهی	ماشین با حالت متناهی	ماشین خودکار با حالت متناهی	√	√
49	کوتاه سازی	کوتاه سازی	کوتاه سازی	√	√

در جدول شماره ۲ در مواردی که در ستون «اصطلاح در متن» بیش از یک واژه یا عبارت دیده می‌شوند این واژگان یا عبارت‌ها به شکل‌های گوناگونی در متن استفاده شده‌اند.

تلاش شده که حتی‌الامکان عین عبارات کتاب، حتی استفاده یا عدم استفاده از فاصله و نیم‌فاصله در واژگان و عبارات در جدول ارائه شود. در این مورد حتی فاصله‌ها نیز در نظر گرفته شده است.

در جدول ۳ سایر اصطلاحات نمایه که جایی برای آن‌ها در جدول ۲ از منظر مقایسه با واژه‌نامه وجود نداشته، آمده‌اند.

جدول ۳: اصطلاحات دیگر نمایه که در جدول ۲ نیامده‌اند

ردیف	سایر اصطلاحات نمایه
۱	ابهام
۲	اجزای کلام
۳	ارزیابی موتورهای جستجوی عربی
۴	اعوجاج مدرک
۵	ان - گرامرها
۶	بازیابی اطلاعات هوشمند نک نظام بازیابی اطلاعات هوشمند
۷	پسوندهای تصریفی
۸	پسوندهای مشتقات
۹	پیش پردازش مدرک
۱۰	تجزیه کننده ها
۱۱	تحلیل تکواژشناختی
۱۲	تحلیل نحوی

تفسیر لفظی	۱۳
دقت بازیافت	۱۴
ریشه‌ساز عربی	۱۵
ریشه‌ساز فارسی	۱۶
ریشه‌سازی	۱۷
زبان اسپانیولی	۱۸
زبان عربی	۱۹
عبارت‌های مرکب	۲۰
طول پرس و جو	۲۱
طول ریشه	۲۲
فهرست واژه‌های تهی	۲۳
فهرست واژه‌های غیرمجاز	۲۴
کنفرانس TREC	۲۵
ماشین‌های ترجمه	۲۶
مجموعه داده‌های Tipster	۲۷
مدل لجستیکی	۲۸
مفاهیم مرکب نک عبارت‌های مرکب	۲۹
مقوله دستوری	۳۰
مقوله‌های مرکب نک عبارت‌های مرکب	۳۱
موتور جستجوی وب عربی	۳۲
میانگین دقت بازیافت	۳۳
نظام بازیابی اطلاعات هوشمند	۳۴
نظام EAGLE	۳۵

در نمایه ترتیب شماره‌های صفحات در برخی مدخل‌ها به هم ریخته است. مدخل‌های «وردنت» و «همجواری» نمونه‌هایی از این مدخل‌ها هستند.

۴. نتیجه‌گیری و چند پیشنهاد

کتاب «پردازش زبان طبیعی و بازیابی اطلاعات» اگر چه کتابی است پیرامون موضوعی مهم و اساسی برای رشته علم اطلاعات و دانش‌شناسی، اما ایراداتی دارد که اگر قرار است این کتاب به چاپ‌های بعدی برسد باید برطرف شوند. اولاً فصل‌ها بهتر است جا به جا شوند: مثلاً فصل دوم به جای فصل اول بیاید زیرا برخی از مفاهیم فصل اول در فصل دوم توضیح

داده شده‌اند. همچنین در فصل هفتم که مربوط به زبان اسپانیایی است تازه مفهوم «ریشه-سازی» توضیح داده شده است. فصل هفتم از جمله فصل‌هایی است که باید در ابتدای کتاب بیاید. محتوای برخی از فصل‌ها برای یک کتاب آموزشی مناسب نیستند. برای نمونه، فصل چهارم صرفاً گزارش یک تجربه است و در حالی که ترجمهٔ چندان مناسبی هم ندارد با هدف کلی این کتاب همخوان نیست. این خصوصیت در مورد فصل ششم بسیار بیشتر صدق می‌کند و این فصل، بسیار فنی و تخصصی است. اگر قرار است این کتاب دوباره چاپ شود حتماً باید توضیحاتی از سوی مترجمان برای روشن‌تر شدن مفاهیم کتاب ارائه شود. سطح ترجمه نیز در فصل‌های مختلف یکسان نیست: مثلاً فصل دوم بهتر از فصل اول ترجمه شده است و این، افزون بر مواردی که در بالا پیرامون ناسلیسی متن ارائه شد، ویراستاری ادبی جدی را می‌طلبد. در صفحات مقدماتی کتاب نیز از هیچ ویراستاری، چه علمی چه ادبی نامی برده نشده است.

کتاب «پردازش زبان طبیعی و بازیابی اطلاعات» کتابی است که اگر قرار است سودمند باشد به ساختاری نو، با ویراستاری علمی همراه با توضیحات فراوان و ویراستاری ادبی جدی نیاز دارد.

۵. یادداشت

دربارهٔ مریم ناصری جستجوی اینترنتی هیچ کمکی نکرد. تنها اطلاعی که کسب شد این بود که وی ویراستار فارسی «مجله بین‌المللی علوم اطلاع‌رسانی و مدیریت اطلاعات» است. نگارندهٔ این سطور حتی پیامی با این محتوا: «سرکار خانم مریم ناصری با سلام و احترام، در حال نوشتن نقدی بر یکی از آثار شما هستم و به اطلاعات مختصری از سرکار نیاز دارم. سپاسگزار میشوم اگر در حد یک پاراگراف اطلاعات اندکی دربارهٔ فعالیتها و آثار علمی خود برای بنده بفرستید. با سپاس فراوان مرتضی کوکبی» برای دو گروه بحث مربوط به رشته فرستاد اما پاسخی دریافت نکرد.

کتاب‌نامه

پردازش زبان طبیعی و بازیابی اطلاعات (۱۳۸۸). ترجمهٔ جعفر مهرداد و مریم ناصری، چاپ اول، تهران: چاپار؛ شیراز: مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری.

۴۱۲ پژوهش‌نامه انتقادی متون و برنامه‌های علوم انسانی، سال بیستم، شماره دهم، دی ۱۳۹۹

جمهوری اسلامی ایران. وزارت علوم، تحقیقات و فناوری. (۱۳۹۲). برنامه درسی مقطع دکترای علم
اطلاعات و دانش‌شناسی: گرایش بازیابی اطلاعات <https://sep.iau.ir/silabes/17416.pdf>

