

*Critical Studies in Texts & Programs of Human Sciences,*  
Institute for Humanities and Cultural Studies (IHCS)  
Monthly Journal, Vol. 20, No. 10, Winter 2020-2021, 367-397  
Doi: 10.30465/crts.2020.5889

## **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: 2020-06-22, 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.



## نقدی بر کتاب پردازش زبان طبیعی و بازیابی اطلاعات

مرتضی کوبی\*

### چکیده

کتاب پردازش زبان طبیعی و بازیابی اطلاعات کتابی است گردآورده و ترجمه‌شده که نویسنده مشخصی ندارد و ترجمه جعفر مهاد و مریم ناصری است. اگر بپذیریم که بازیابی قلب رشته علم اطلاعات و دانش‌شناسی است، می‌توانیم بگوییم که اگر این بازیابی با زبان طبیعی انجام شود، نتیجه برای آن سودمندتر خواهد بود. پس کتابی که درباره «پردازش زبان طبیعی و بازیابی اطلاعات» باشد، می‌تواند برای رشته و حرفه از ارزش بالایی برخوردار باشد. افزون‌براین، در گرایش «بازیابی اطلاعات و دانش» دوره دکترای علم اطلاعات و دانش‌شناسی، از ۱۳۹۱ درسی تخصصی الزامی با عنوان «پردازش زبان طبیعی» ارائه شده که باتوجه به کمبود (و شاید هم نبود) منابعی برای این درس وجود کتابی با محتوای پردازش زبان طبیعی می‌تواند فی‌نفسه سودمند باشد؛ اما کتاب ایرادهایی دارد که در تحلیل و ارزیابی محتوایی اثر با تفصیل بیش‌تری درباره آن‌ها بحث خواهد شد. درباره رعایت اصول علمی ارجاع‌دهی در درون متن و کتاب‌نامه (که در پایان هر فصل آمده است) نیز در «تحلیل و ارزیابی محتوایی اثر» مطالبی ارائه خواهد شد. رابطه میان اصطلاحات ترجمه‌شده در متن، واژه‌نامه، و نمایه در کتاب موردنقد با استفاده از جدول 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، ایران، Kokabi80@yahoo.com

تاریخ دریافت: ۱۳۹۹/۰۴/۰۲، تاریخ پذیرش: ۱۳۹۹/۰۸/۱۰

## ۱. مقدمه

عبارت «پردازش زبان طبیعی» به‌منزلهٔ درسی از درس‌های گرایش «بازیابی اطلاعات» دورهٔ دکترای علم اطلاعات و دانش‌شناسی شاید نخستین‌بار به‌طور رسمی در مصاحبهٔ جعفر مهرداد، سرپرست پیشین کمیتهٔ برنامه‌ریزی علم اطلاعات و دانش‌شناسی وزارت علوم، تحقیقات و فناوری در مصاحبه با خبرگزاری *لینا* اعلام شد. بخش‌هایی از این مصاحبه عیناً در ادامه آورده شده است:

جعفر مهرداد، سرپرست کمیتهٔ برنامه‌ریزی علم اطلاعات و دانش‌شناسی وزارت علوم در گفت‌وگو با *لینا* با اعلام خبر تصویب اولین برنامهٔ دورهٔ دکترای علم اطلاعات و دانش‌شناسی اظهار داشت: 'از سال ۱۳۸۶ تاکنون برنامه‌های مختلفی در مقاطع مختلف تحصیلی توسط این کمیته برنامه‌ریزی شده و به تصویب رسیده که مصوبات کمیته رسماً توسط وزارت علوم، تحقیقات و فناوری به دانشگاه‌ها ابلاغ شده است. وی افزود: 'در جلسهٔ ۱۳۹۱/۱۲/۱۵ بعد از چندین ماه برنامه‌ریزی دربارهٔ گرایش بازیابی اطلاعات و دانش در مقطع دکترا، لاجرم این گرایش که نخستین برنامهٔ دورهٔ دکترای علم اطلاعات و دانش‌شناسی است به تصویب رسید.'

به‌گفتهٔ مهرداد، تعداد درس‌های پایهٔ این گرایش چهار درس به شرح زیر است:

۱. فلسفه علم؛
  ۲. نظریه‌های اطلاعات و دانش؛
  ۳. طبقه‌بندی و توسعهٔ علوم؛
  ۴. سیبرنتیک و ارتباطات.
- وی تصریح کرد: 'برای این گرایش ۶ درس تخصصی الزامی نیز به‌تعداد ۱۲ واحد منظور شده است:

۱. پردازش زبان طبیعی؛
۲. وب معنایی و هستی‌شناسی؛
۳. تحلیل متن و دانش‌کاوی؛
۴. معماری اطلاعات؛
۵. ارزیابی نظام‌های بازیابی اطلاعات و دانش؛
۶. سمینار مسائل ویژه در بازیابی اطلاعات و دانش ' (برنامهٔ درسی مقطع دکترای علم

همان گونه که مشاهده می‌شود، درس «پردازش زبان طبیعی» نخستین درس از شش درس تخصصی الزامی در گرایش «بازیابی اطلاعات و دانش» از دوره دکترای علم اطلاعات و دانش‌شناسی است. قرارداد 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<sup>۱</sup> 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 قدیم‌ترین استادان رشته علم اطلاعات و دانش‌شناسی کشور و صاحب کتب بسیاری است. او در سمت‌های اجرایی رده‌بالایی نیز فعالیت داشته و در کل می‌توان او را جزو نامورترین استادان رشته به‌شمار آورد.

کتاب هشت فصل به شرح ذیل دارد:

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: جست‌وجوی 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 افزون‌بر این هشت فصل، یک واژه‌نامه انگلیسی - فارسی و یک نمایه نیز دارد.

اما کتاب ایرادهایی دارد که در تحلیل و ارزیابی محتوایی اثر با تفصیل بیش‌تری درباره آن‌ها بحث خواهد شد. درباره رعایت اصول علمی ارجاع‌دهی در درون متن و کتاب‌نامه که در پایان هر فصل آمده است نیز در تحلیل و ارزیابی محتوایی اثر مطالبی ارائه خواهند شد. رابطه میان اصطلاحات ترجمه شده در متن، واژه‌نامه، و نمایه در کتاب موردنقد با استفاده از جدول 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 شیوه‌سندهی و ارجاع نقد و ارزش‌یابی می‌شود. در نقد محتوایی، محتوا از منظرهای علمی، اخلاقی، دینی، تاریخی، و غیره نقد می‌شود. بدیهی است کتابی را که درباره‌ی پردازش زبان طبیعی و بازیابی اطلاعات است نمی‌توان از منظرهای اخلاقی و دینی نقد کرد، اما آنچه در نقد حاضر بیش‌تر موردتوجه 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، رنگ‌های گرم (نارنجی و قرمز) دارد، اما تصویر روی آن معلوم نیست که نشان‌دهنده‌ی چیست (حداقل نگارنده‌ی این سطرها از آن چیزی نفهمیده است). صحافی، قطع، نوع و اندازه‌ی 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 رشته‌ی علم اطلاعات و دانش‌شناسی از اهمیت زیادی برخوردار است. اگر بپذیریم که بازیابی قلب رشته‌ی علم اطلاعات و دانش‌شناسی است، آن‌گاه می‌توانیم بگوییم که اگر این بازیابی با زبان طبیعی انجام شود، زبانی که کاربر با آن زندگی، و به تبع اطلاعات را با آن جست‌وجو می‌کند، نتیجه برای او بسیار سودمندتر خواهد بود. پس کتابی که درباره‌ی «پردازش زبان طبیعی و بازیابی اطلاعات» باشد، می‌تواند برای رشته و حرفه از ارزش بالایی برخوردار باشد. افزون‌براین، به این نکته اشاره شد که در گرایش «بازیابی اطلاعات و دانش» دوره‌ی دکترای علم اطلاعات و دانش‌شناسی، از سال ۱۳۹۱ درسی تخصصی الزامی با عنوان «پردازش زبان طبیعی» ارائه شده، که باتوجه 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 این‌که 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 به منزله متنی درسی برای این درس انتخاب شود. یکی از مترجمان، جعفر مهرداد، از استادان قدیمی رشته است که از نیاز رشته و حرفه به این کتاب نیک آگاه است 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 نکته‌های مثبت کتاب است.

از آنجایی که هدف نقد بیان اشکالات در کنار محاسن اثر است، آنچه در ادامه می‌آید نگاهی به محتوای فصل‌های گردآورده و ترجمه شده است. در پیوند با گردآوری و ترجمه یکی از نکته‌هایی که در کتاب به چشم می‌خورد، استفاده از واژه «ترجمه» فقط در روی جلد و صفحه عنوان کتاب است، در حالی که در فهرست نویسی پیش از انتشار برای مترجمان از واژگان «گردآورنده و مترجم» استفاده شده 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 نشده است. 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 هیچ‌جای کتاب پاسخ داده نشده است. حتی در مقدمه کتاب که به قلم استاد قدیمی نوشته شده و نیازمند ویرایش ادبی اندکی است، هیچ اشاره‌ای به این نکته نشده است. فقط در صفحه سوم از مقدمه چنین نوشته شده است: «مطالبی که در فصول مختلف این کتاب در پیش‌رو دارید گزیده‌ای است آگاهانه از مطالعاتی که در مراکز علمی جهان در حیطه زبان [زبان‌های] فارسی، اسپانیولی و انگلیسی صورت گرفته است». اما درباره 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 سطر ۲ از صفحه 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 سطر ۱ از صفحه ۴ نیز عبارت «مثال‌های ۴-۱ پیش‌گفته را مدیریت می‌کند» دیده می‌شود، اما این مثال‌ها در این فصل دیده نمی‌شوند. مجموعه این عوامل نشان می‌دهد که کتاب گردآورده از چند منبع است که به دلایلی معرفی نشده‌اند. در سطر اول از صفحه ۳۲ (فصل سوم) نیز واژه «فصل» دیده می‌شود. در همین فصل، در عنوان‌های برخی شکل‌ها

به شکل‌ها و جدول‌هایی ارجاع داده شده است که در این فصل دیده نمی‌شوند. در صفحه ۵۴ نیز واژه «فصل» دیده می‌شود. علت (علل) و معیار(های) این محذوفات اگر وجود دارد چه بوده است؟

همان‌گونه که پیش‌تر اشاره شد، ترتیب فصل‌ها نیز جای بحث دارد: مثلاً معلوم نیست که چرا مطالب مربوط به زبان‌های فارسی و عربی در آخر آمده‌اند، درحالی‌که مطالب مربوط به زبان اسپانیایی (که معلوم نیست چرا اصلاً انتخاب شده) میان این دو آمده است. ابهام درباره 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 است که مقدمه کتاب به ویرایش ادبی اندکی نیاز دارد. نمونه‌ای از این نیاز به ویرایش 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 مقدمه جمله‌ای وجود دارد که معنای آن چندان رسا نیست: «به‌خاطر توجه به علایق خوانندگان این اثر از اصطلاح قدیمی‌تر به‌عنوان بازیابی مدرک و از اصطلاح جدیدتر به‌عنوان بازیابی متن به‌طور 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، درحالی‌که این دو متفاوت‌اند. در سطرهای ۱۴ تا ۱۸ صفحه «و» مقدمه نیز این پاراگراف دیده می‌شود:

توصیف‌گرها که ممکن است غیرمستقیم باشند، چون متون کامل مدارک که به‌طور فزاینده‌ای [به‌صورت روی خط در دسترس می‌باشند، موضوع توصیف‌گرها در مقایسه با سال‌های گذشته بیش‌تر تحت‌کنترل روش‌های نمایه‌سازی است. باوجوداین، همیشه نمی‌توان توصیف‌گرها را به‌طور کامل تهیه دید. نویسنده یک مدرک همیشه گفته‌های فراوانی دارد که در مدرک به‌طور 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 باید در ویراست‌های بعدی کتاب برطرف شوند.

نیاز به ویرایش ادبی در جاهای دیگری از کتاب هم دیده می‌شود و تعداد این موارد کم هم نیست. این موارد شامل اصطلاحاتی‌اند که درست معادل‌یابی نشده یا جمله‌هایی هستند که یا درست ترجمه نشده یا تحت‌اللفظی ترجمه شده‌اند. یکی از این موارد عبارت «نظام بازیابی اطلاعات هوشمند» است؛ از آنجایی 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 سراسر متن رعایت شود. این نکته در مورد «بازیابی اطلاعات قراردادی» نیز صادق است. در پاراگراف آخر صفحه ۱۷ نیز عبارت «پردازش زبان طبیعی خودکار» باید «پردازش خودکار زبان طبیعی» باشد. در سطر اول صفحه ۱۸، عبارت «اختلاف زبانی» باید «گونه‌گونی زبان‌شناختی» ترجمه شود؛ معادل انگلیسی این اصطلاح «linguistic variation» است. همین‌طور است ترجمه پانویس شماره ۱ در صفحه 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 صفحه ۴۸ بهتر است «اسم فعل» یا «اسم مصدر» ترجمه شود. در پانویس ۱ صفحه ۷۵ «bayes theorem» باید «bayes theorem» باشد. در صفحه ۴ سطر ۱۵ تا ۱۶ نمونه‌ای از ترجمه ناسلیس دیده می‌شود: «باوجود این، بخش تاثیر [تأثیر] ریشه‌سازی در بازیابی ناچیز نمی‌باشد». جمله «امتیازی که جداسازی مفاهیم مرکب دارد مثبت می‌باشد» در سطر ۲ از صفحه ۸ نیز از همین قبیل است. در صفحه ۹ نیز با این جمله روبه‌رو می‌شویم: «وورهایس (۱۹۹۳) دریافت وقتی که از داده‌های انگلیسی با این روش استفاده می‌شود دقت ۳ امتیازی بین ۵ تا ۱۷٪ کاهش می‌یابد». جمله پس از این جمله هم اشکال دارد: «وُلک و همکاران او (۲۰۰۲) آزمون مشابهی را بر روی داده‌های آلمانی و انگلیسی انجام داد[ند] و نتایج به‌هم‌آمیخته‌ای را گزارش نمود[ند]». منظور از «به‌هم‌آمیخته» در این جمله روشن نیست. در سطر ۲ از صفحه ۱۲ ترکیب «مزایای خوبی» دیده می‌شود: آیا مزایای بد هم وجود دارند؟! در سطر ۱۳ از

صفحه ۳۴، می‌توان به‌جای استفاده از واژه نامناسب «با اثر»، واژه «مؤثر» را به‌کار برد. در سطر ۴ از صفحه ۳۵، «انفجار پایگاه‌های اطلاعاتی غیرکتاب‌شناختی شایع» هم از آن دسته جملاتی است که درباره آن باید بسیار اندیشید! مفهوم این جمله در صفحه ۳۱ روشن نیست: «ترویج بین نظام بازیابی اطلاعات و فنون هوش مصنوعی، طبیعی و به‌صورت بالقوه با اهمیت [بااهمیت] می‌باشد». منظور از واژه «توسعه‌های» در عبارت «توسعه‌های گوناگونی نسبت به فناوری نمایه معکوس استاندارد پیش‌نهاد شده است» در سطر ۳ از صفحه ۳۳ مبهم است. در همین صفحه، عبارت «inverted index» بهتر است به‌جای «نمایه معکوس»، «نمایه مقلوب» ترجمه شود. در صفحه ۳۶ نیز به این جملات برمی‌خوریم که درک آن‌ها دشوار است:

«باوجود این، اسناد الزاماً به‌شکل زبان طبیعی یا متن آزاد که بازیابی اطلاعات هوشمند [بازیابی هوشمند اطلاعات] برای استنتاج و پی‌بردن به توان بازیابی اطلاعات نیاز دارد موجود و قابل استفاده می‌باشند».

«در محیط بازیابی اطلاعات هوشمند [بازیابی هوشمند اطلاعات] مؤلفه‌ای [مؤلفه‌ای] زبان‌شناختی می‌تواند روابط اصطلاح را بین واحدهای نمایه‌سازی که ممکن است برای تولید توصیف‌های کامل‌تر مدرک در برابر توصیف حاصل از اصطلاحات واحد به‌کار برد نشان دهد».

ترویج بین نظام بازیابی اطلاعات و فنون هوش مصنوعی که شامل پردازش زبان طبیعی، بازنمون دانش، استدلال و استنتاج است طبیعی و بالقوه مفید و با اهمیت [بااهمیت] است». در سطر ۱۷ تا ۱۸ از صفحه ۴۲ نیز جمله زیر نشان از نبود سلاست در ترجمه دارد: «هم‌چنین، از EAGLE برای نرمال‌سازی اشکال تک‌واژشناختی واژه‌ها استفاده کردیم، طوری که واژه «foot» با «feet» تطبیق خواهد کرد». مفهوم «تطبیق» در این جمله مبهم است. هم‌چنین در سطر ۵ از صفحه ۴۳ جمله «برای رعایت کردن پرس‌وجوهای کوتاه آن‌گونه که در جست‌وجوی وب به‌کار می‌روند، آزمون نظام خود را در محیط واقعی انجام دادیم» نامفهوم است. همین‌گونه هستند سطرهای ۲ تا ۴ از آخر در همین صفحه که عبارت‌اند از: «از این طریق، برای هر پرس‌وجو تعداد ۱۸۰ سند به‌دست آوریم [آوردیم] که درباره نتایج حاصل از یک موتور کاوش وب نیز مصداق دارد». در مورد شماره ۴ در صفحه ۴۸، در جمله «۴. ریشه‌های واژه را تنها با نادیده‌گرفتن مورفولوژی و برجسب‌های اجزای کلام تطبیق دهید» معلوم نیست که ریشه‌ها را باید با چه تطبیق داد. در ضمن، واژه مورفولوژی را می‌توان «ریخت‌شناسی» ترجمه کرد. این واژه «تطبیق» در موارد ۱ و ۵ در همین صفحه نیز

ابهام‌برانگیز است. درک مفهوم این جمله در سطرهای ۳ تا ۴ از صفحه 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 سطر ۴ تا 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 صفحه 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 کارآیی دارد که پاراگراف مطرح در آن به انگلیسی آورده شده باشد. هم‌چنین معلوم نیست چرا جدول‌ها و شکل‌های فصل سوم ترجمه نشده‌اند؛ این‌ها متن‌های ساده‌ای دارند 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 به حالت سوم شخص مفرد باشند. اشتباهات این جدول جای بحث فراوانی دارد. در جدول ۲ در صفحه ۵۵، در حالی فعل «کردن» به منزله 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 این‌که مثال «کردن» در جدول ۲، ناقض 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 صفحه ۵۹، به ردیف «S» و ستون «I» اشاره شده، اما این حرف و عدد در هیچ‌یک از تصاویر دیده نمی‌شوند. در سطر ۱۴ تا ۱۷ صفحه ۷۹ این جمله دیده می‌شود: «بنابراین، قراردادن سیستمی از یک سازوکار [سازوکار یا ساز و کار] که استاندارد کردن واژه‌های مختلف را امکان‌پذیر سازد و بدین ترتیب اشکال مختلف یک ریشه واحد را زیر یک شکل واحد که ممکن است در پرس و جو و در مدرک وجود داشته باشند [باشد] ضروری می‌باشد». این جمله هم نامفهوم و هم ناسلیس است. در صفحه ۳۴ کتاب نیز در انتهای پاراگراف اول ادعایی مطرح شده است 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 صفحه ۴۶ از مفهوم «پنجره سه‌جمله‌ای» استفاده شده که نامفهوم و قبلاً درباره آن توضیحی داده نشده است. در سطر ۱۸ تا ۱۹ از صفحه 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 کروتز در صفحه 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 صفحه ۱۴ دارای اشکال و شکلی غیرعادی است. ضمن این‌که در مواردی به نامی در متن استناد شده که در فهرست منابع آن فصل وجود ندارد: «هال و همکاران (۱۹۹۶)» در صفحه ۸ نمونه‌ای از این موارد است. در مواردی استناد درون‌متنی با شناسه مربوط در ارجاعات هم‌خوانی ندارد؛ مثال: (Manning 1999). در صفحه ۲۱ که معلوم نیست چرا به انگلیسی نوشته شده در ارجاعات دارای نویسنده همکار است که در استناد درون‌متنی 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 صفحه ۹۹، ظاهراً دو منبع است که تجانسی باهم ندارند. این اشکالات باید بازننگری و اصلاح شوند.

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

#### ۴. برخی ایرادهای جزئی

ایرادهای جزئی هم در کتاب دیده می‌شود؛ در صفحه ۴ دو پانویس وجود دارند که شماره‌های یک‌سانی دارند. اشتباهات تایپی هم در زمره این ایرادهای جزئی هستند: در

شکل ۱ در فصل دوم، واژه «cunction» باید «function» باشد. در صفحه ۴۵، فرمولی برای «relevance» (ربط) وجود دارد که در آن، ظاهراً واژه «token» (به معنای نماد) چهار بار به یکی از صورت‌های زیر اشتباه نوشته شده است: «tokanes, taken, takanes». واژه «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 سراسر کتاب استفاده شود. برخی از ارقام درون قلاب در فصل چهارم مانند آنچه در صفحه ۶۰ آمده است، با شماره‌های مربوط به ارجاعات اشتباه می‌شوند. تلفظ برخی نام‌های خارجی نیز اشتباه است؛ نام «بوکستاین» در صفحه ۷۵ باید «بوکستاین» باشد.

در واژه‌نامه که از نکته‌های مثبت کتاب است ۳۳ اشتباه وجود دارد. برخی ناهماهنگی‌ها در ترجمه نیز در واژه‌نامه مشخص شده‌اند: در واژه‌نامه «dictionary-based» «فرهنگ‌محور» ترجمه شده است که درست هم نیست؛ درعین حال «memory-based» «حافظه‌بنیاد» ترجمه شده؛ همین‌طور «rule-based» که «قاعده‌بنیاد» است.

## ۵. رابطه میان اصطلاحات ترجمه‌شده در متن، واژه‌نامه و نمایه

رابطه میان اصطلاحات ترجمه‌شده 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 درخور توجه است. این رابطه از طریق دو جدول زیر نمایش داده می‌شود. جدول ۱ رابطه میان اصطلاحات ترجمه‌شده در متن، واژه‌نامه، و نمایه را نشان می‌دهد.

جدول ۱. رابطه میان اصطلاحات ترجمه‌شده 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	برابری تعداد واژگان در واژه‌نامه و نمایه
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	✓	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 نک مقوله‌های 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 بازآور	-	
۱۸	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	تجزیه گر تابع	Dependency Parser	Dependency Parser	√	تجزیه گر تابع	√	
۲۳	ابهام زدایی از مفهوم واژه، ابهام زدایی از معانی واژه‌ها	Word-Sense Disambiguation	Word-sense Disambiguation	√	ابهام زدایی از مفهوم واژه	ابهام زدایی از معانی واژه‌ها	
۲۴	تجزیه گر حالت نامتناهی**	Finite State Parser	Finite State Parser	√	تجزیه گر 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	Pragmatic Level	√	سطح منظور شناختی		سطح منظور شناختی
۳۹	تفسیر لفظی	Literal Interpretation	Literal Interpretation	√	تفسیر لفظی	√	تفسیر لفظی
۴۰	ربط	Relevance	Relevance	√	ربط	√	ربط
۴۱	نمایه اصطلاحات	Terms Index	Terms Index	√	نمایه اصطلاحات	√	-



۴۲	بسته‌ای از واژه‌ها	Bag of Words	Bag of Words	√	بسته ای از واژه‌ها	-
۴۳	مولفه‌ها	Parameters	Parameters	√	مولفه	-
۴۴	اصطلاحات مربوط	Relevant Terms	Relevant Terms	√	اصطلاحات مرتبط	-
۴۵	سرآیندها	Header	Headers		سرآیند	-
۴۶	سرنام‌ها	Acronyms	Acronyms	√	سرنام	-
۴۷	حروف اضافه	Preposition	Prepositions		حروف اضافه	√
۴۸	شکل‌های تصریفی	Inflected Form	Inflected Forms		شکل تصریفی	-
۴۹	فراوانی مدرک، فراوانی مدرک معکوس	Term Frequency (TF)	Term Frequency, Inverse Document Frequency		فراوانی مدرک	-
۵۰	ارزش تمیز دهنده	Discriminatory Value	Discriminatory Value	√	ارزش تمیز دهنده	√
۵۱	فهرست واژه‌های غیرمجاز	Stop word List	Stopwords [Stop words] List		واژه غیر مجاز	واژه‌های غیرمجاز
۵۲	تجزیه کننده‌ها	Parser	Parsers		تجزیه کننده	تجزیه کننده ها
۵۳	جمله‌های اسمی	Nominal Sentences	Nominal Sentences	√	جملات اسمی	-
۵۴	جملات اضافی	Prepositional Sentences	Prepositional Sentences	√	جملات اضافی	√
۵۵	پایگاه اطلاعاتی واژگانی وُردنت	Lexicographic Database WordNet	Lexicographic Database WordNet	√	پایگاه اطلاعات واژگانی وُردنت	وردنت
۵۶	مجموعه‌های مترادف	Synsets	Synsets	√	مجموعه‌های مترادف	-
۵۷	موجودیت‌ها	Entities	Entities	√	موجودیت	موجودیتها

۵۸	پایگاه اطلاعاتی هستی‌شناسی		هستی‌شناسی	√	Ontology Database	Ontology Database	پایگاه اطلاعاتی هستی‌شناسی	
۵۹	اسمارت		اسمارت (زبان طبیعی اسمارت)	√	SMART	SMART	نظام زبان طبیعی اسمارت	
۶۰	√	میان‌زبانی	√	میان‌زبانی	√	Interlingual	Interlingual	میان‌زبانی
۶۱	-	-	-	-	-	Empty Phrases	عبارت‌های تهی	
۶۲	-	√	قراردادی	√	Conventional	Conventional	قراردادی	
۶۳	√	هوش مصنوعی	√	هوش مصنوعی	√	Artificial Intelligence	Artificial Intelligence	هوش مصنوعی
۶۴	√		√	فوق‌متن		Hyper Text	Hypertext	فوق‌متن
۶۵	√	بردارهای اصطلاح		بردار اصطلاح		Term Vector	Term Vectors	بردارهای اصطلاح
۶۶	√	اپراتورهای بولی	√	اپراتورهای بولی		Boolean Operator	Boolean Operators	اپراتورهای بولی
۶۷	√	نمایه معکوس	√	نمایه معکوس	√	Inverted Index	Inverted Index	نمایه معکوس
۶۸	-	√	زبان طبیعی	√	Natural Language	Natural Language	زبان طبیعی	
۶۹	√	متن آزاد	√	متن آزاد		Free Text	Free-text	متن آزاد
۷۰	√	واژه‌های هم‌نویسه	√	واژه‌های هم‌نویسه	√	Homographs	Homographs	واژه‌های هم‌نویسه
۷۱	-		√	تفسیر منظورشناختی	√	Pragmatic Interpretation	Pragmatic Interpretation	تفسیر منظور شناختی
۷۲		شناسگرها		شناساگر	√	Recognizer	Recognizer	شناسگر
۷۳		رتبه‌بندها		رتبه‌بند	√	Ranker	Ranker	رتبه‌بند
۷۴	-			ماجول	√	Module	Module	ماجول‌ها

نقدی بر کتاب پردازش زبان طبیعی و بازیابی اطلاعات (مرتضی کوبی) ۳۸۵

۷۵	رشته	String	String	رشته	√	رشته	√	-
۷۶	عبارت اسمی	Noun-Phrase	Noun-phrase	عبارت اسمی	√	عبارت اسمی	√	-
۷۷	شمول معنایی	Hyponymy	Hyponymy	شمول معنایی	√	شمول معنایی	√	شمول معنایی
۷۸	بسته وُرد نِت پرنیستون [پرنیستون]	Princeton Word Net Package	Princeton Word Net Package	بسته وردنت پرنیستون	√	بسته وردنت پرنیستون	√	-
۷۹	زیر مفهوم	Sub Concept	Sub-concept	زیر مفهومی	√	زیر مفهومی	√	-
۸۰	آلتا ویستا	Alta Vista	Alta Vista	آلتا ویستا	√	آلتا ویستا	√	آلتا ویستا
۸۱	فضای نمادی	Token Space	Token Space	فضای نمادی	√	فضای نمادی	√	-
۸۲	نسخه "وانیلا"ی فضای برداری	Vanila Version of Vector Space	Vanila Version of Vector Space	نسخه وانیلای فضای برداری	√	نسخه وانیلای فضای برداری	√	-
۸۳	همجواری	Proximity	Proximity	همجواری	√	همجواری	√	همجواری
۸۴	جفت جلویی	Forward Paris [Pairs]	Forward Pairs	جفت جلویی	√	جفت جلویی	√	جفت جلویی
۸۵	جفت عقبی	Backward Paris [Pairs]	Backward Pairs	جفت عقبی	√	جفت عقبی	√	جفت عقبی
۸۶	ثبات‌ها	Register	Registers	ثبت	√	ثبت	√	-
۸۷	پشتیبانی	Backing off	Backing Off	پشتیبانی	√	پشتیبانی	√	-
۸۸	مصدرهای فعلی	Gerund	Gerunds	مصدر فعلی	√	مصدر فعلی	√	-
۸۹	شناسگر عبارت اسمی	Noun Phrase Recognizer	Noun Phrase Recognizer	شناساگر عبارت اسمی	√	شناساگر عبارت اسمی	√	شناسگرها
۹۰	فارسی	-	Farsi	-	√	-	√	-

۹۱	ریشه‌ساز لاوینز	Lovins Stemmer	ریشه‌ساز لاوین	-	-
۹۲	ریشه‌ساز پورتر	Porter Stemme [Stemmer]	ریشه‌ساز پورتر	ریشه‌ساز پورتر	√
۹۳	دقت	Precision	دقت	√	دقت
۹۴	بازیافت	Recall	بازیافت	√	بازیافت
۹۵	پسوندها	Suffixes	پسوندها	-	-
۹۶	پیشوندها	Prefixes	پیشوندها	-	-
۹۷	نفی	Negation	نفی	√	-
۹۸	وجه	Mood	وجه	√	-
۹۰	مصدر	Infinitive	مصدر	√	-
۹۱	مضارع اخباری	Indicative Present Tense	مضارع اخباری	√	-
۹۲	تفضیلی	Comparative	تفضیلی	√	-
۹۳	مصمت‌ها	Consonants	مصمت	-	-
۹۴	مصوت‌ها	Vowels	مصوت	-	-
۹۵	زیر رشته انتهایی	Terminal Substring	زیر رشته انتهایی	√	-
۹۶	حالت اضافه جمع	Plural Possessive	حالت اضافه جمع	√	-
۹۷	پسوندهای فعلی	Verb- suffixes	پسوندهای فعلی	-	-
۹۸	پسوندهای اسم جمع	Plural-noun-suffixes	پسوندهای اسم جمع	-	-
۹۹	پسوندهای اسم ملکی	Possessive- noun-suffixes	پسوندهای اسم ملکی	-	-

۱۰۰	ماشین خودکار حالت قطعی با نامتناهی	√	ماشین خودکار قطعی با حالت متناهی		Deterministic Finite Automata	Deterministic Finite Automata (DFA)	ماشین خودکار قطعی با حالت متناهی
۱۰۱	پس-پردازشگر		پس پردازشگر	√	Post Processor	Post Processor	پس-پردازشگر
۱۰۲	-		بردار محور	√	Vector-based	Vector-based	بردار-محور
۱۰۳	یورو فارسی		یورو فارسی	√	Euro Farsi	Euro Farsi	یوروفارسی
۱۰۴	-	√	قطعه‌ای کوچک از کد که بر روی اینترنت قابل انتقال باشد و بر روی ماشین دریافت کننده اجرا شود.	√	Applet	Applet	قطعه‌ای کوچک از کد که بر روی اینترنت قابل انتقال باشد و بر روی ماشین دریافت کننده اجرا شود.
۱۰۵	-		مجموعه‌ای متوالی از بیت‌هایی که تصویری را در حافظه معرفی می کند 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 می کنند که باید بر روی صفحه نمایش داده شود، به ویژه در سیستم‌هایی که رابط گرافیکی کار بر دارند.

	-		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 صفحات وب اضافه کرد.	√	Java Script	Java Script	زیر مجموعه‌ای از زبان جاوا در مقایسه با جاوا عملکرد ضعیفی دارد. برنامه‌های کاربردی فوری، اصلی و توابع را می‌توان با سند اصلی جاوا به صفحات وب اضافه کرد.	۱۱۰

	-	√	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.	√	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

۱۱۸	حاصلضرب داخلی	Inner Product	Inner Product	√	حاصلضرب داخلی	√	-
۱۱۹	قضیه بیزی	Bayes Theorem [Theorem]	Bayes Theorem [Theorem]	√	قضیه بیزی	√	-
۱۲۰	مدل سازی زبان آماری	Statistical Language Modelling	Statistical Language Modeling		مدل سازی زبان آماری	√	-
۱۲۱	توزیع پواسون منحنی ریاضی که غالباً در آمار و شبیه سازی از آن استفاده می شود تا صحت وقوع رویدادها را نشان دهد.	Poisson Distribution	Poisson Distribution	√	توزیع پواسون: منحنی ریاضی که غالباً در آمار و شبیه سازی از آن استفاده می شود تا صحت وقوع رویدادها را نشان دهد.		-
۱۲۲	فنون نرم و روان	Smoothing Techniques	Smoothing Techniques	√	فنون نرم روان		
۱۲۳	الگوریتم داوسون	Dawson	Dawson	√	الگوریتم داوسون	√	-
۱۲۴	الگوریتم پایس	Paice Stemmer	Paice		ریشه‌ساز پایس		-
۱۲۵	واج گونه	Allomorph	Allomorph	√	واج گونه		-
۱۲۶	حروف میانه	Intermediate Letters	Intermediate Letters	√	حروف میانه	√	-
۱۲۷	مسندهای فعلی	Verbal Predicates	Verbal Predicates	√	مسندهای فعلی		-
۱۲۸	مصوت بی تکیه	Unstressed vowel	Unstressed Vowel		مصوت بی تکیه	√	-
۱۲۹	غیر-آسکی	Non-ASCII	Non - ASCII		غیر آسکی		
۱۳۰	سرویس دهنده - پردازش	Server Processing	Server - Processing		سرویس دهنده پردازش		



			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 استانداردی که امکان انتقال داده هایی مانند تصویر، صدا و فایل های دودویی را توسط پست الکترونیکی شبکه اینترنت بدون هیچ اجباری برای تبدیل آن با فرمت ASCII فراهم می سازد.	۱۳۱
√	کوتاه سازی	√	کوتاه سازی	√	Truncation	Truncation	کوتاه سازی	۱۳۲
	-	√	ترتیب واژه با شکل آزاد		Free Word Order	Free word-order	ترتیب واژه با شکل آزاد	۱۳۳
	-	√	اشتقاقی	√	Derivative	Derivative	اشتقاقی	۱۳۴
	-	√	واژه جامد	√	Primitive	Primitive	واژه جامد	۱۳۵
	-		-		-	Participles	اسامی فاعلی و مفعولی	۱۳۶
	-	√	ضمایر موصولی		Reletives [Relatives]	Relatives	ضمایر موصولی	۱۳۷
	-	√	ضمایر اشاره	√	Demonstrative	Demonstrative	ضمیر اشاره	۱۳۸

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): واژه‌هایی مانند gosh! ، ugh! یا Wow که بیانگر نوعی حالت عاطفی یا نگرشی می‌باشد	Interjection	Interjection	√	واژه‌های تعجبی مانند Wow! ، gosh! یا uyh! که بیانگر نوعی حالت عاطفی یا نگرشی می‌باشد	-
۱۴۵	پایه کلمه	Root	Root	√	پایه کلمه	-
۱۴۶	ریشه کلمه	Stem	Stem	√	ریشه کلمه	-
۱۴۷	مالکیت مدرک	Document Privacy	Document Privacy	√	مالکیت مدرک	-
۱۴۸	آینَ	Ayna	Ayna	√	آینَ	-
جمع	۱۴۸	۱۴۴				

\* در نمایه «مقولۀ دستوری» وجود دارد.

\*\* این عبارت باید «تجزیه‌گر حالت متناهی» باشد. درکل، برخی از معادل‌ها اشتباه به نظر می‌رسند.

در جدول بالا، واژگان اشتباه در قلاب اصلاح شده‌اند. نابرابرها در جدول بالا (آنهایی که دارای نشانه  $\surd$  نیستند) شامل نبود معادل، تفاوت در فاصله‌های پیش و پس از حروف و نشانه‌ها، تفاوت در جمع و مفردبودن واژگان یا عبارات، بود یا نبود خط تیره میان دو واژه یا عبارت، اشتباهات تایپی که به‌نوعی باعث تفاوت شده‌اند، بزرگی و کوچکی حروف آغازین در واژگان یا عبارات‌های انگلیسی، چسبیده‌بودن 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 درصد) در واژه‌نامه عیناً با معادل انگلیسی مربوط در متن، مشابهند. معادل‌های فارسی ۸۶ واژه یا عبارت (حدود ۵۸/۱ درصد) 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 درصد از واژگان و عبارات واژه‌نامه را تشکیل می‌دهند.

جدول ۲ رابطه میان اصطلاحات ترجمه‌شده 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 ترجمه‌شده 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
۱	دقت بازیافت با ۱۱ امتیاز؛ دقت با میانگین ۱۱ امتیاز	دقت بازیافت با ۱۱ امتیاز	دقت بازیافت با ۱۱ امتیاز	$\surd$	
۲	دقت بازیافت R	دقت بازیافت R	دقت بازیافت R	$\surd$	$\surd$
۳	مدل فضای برداری	مدل فضای برداری	مدل فضای برداری	$\surd$	$\surd$
۴	ریشه‌سازی تصریفی	-	ریشه‌سازی تصریفی		
۵	ریشه‌سازی اشتقاقی	ریشه‌سازی اشتقاقی	ریشه‌سازی اشتقاقی		
۶	مقوله نحوی؛ مقوله دستوری	مقوله نحوی	مقوله نحوی نک مقوله های دستوری*		
۷	اسمارت	اسمارت (زبان طبیعی اسمارت)	اسمارت		$\surd$

۸	تقطیع واژه‌ها	تقطیع واژه	تقطیع واژه‌ها	√	
۹	جفت‌های سر - تعدیل‌گر	جفت‌های سر تعدیل‌گر	جفت‌های سر - تعدیل‌گر		
۱۰	ابهام زدایی از مفهوم واژه، ابهام زدایی از معانی واژه‌ها	ابهام زدایی از مفهوم واژه	ابهام زدایی از واژه‌ها		
۱۱	چند معنایی	چند معنایی	چند معنایی	√	√
۱۲	هم معنایی	هم معنایی	هم معنایی	√	√
۱۳	اعوجاج مدرک	اعوجاج مدرک	اعوجاج مدرک	√	√
۱۴	سطح تکواژ شناختی	سطح تکواژ شناختی	سطح تکواژ شناختی	√	
۱۵	سطح نحوی	سطح نحوی	سطح نحوی	√	√
۱۶	اختلاف واژگان	اختلاف واژگان	اختلاف واژگان	√	√
۱۷	سطح منظور شناختی	سطح منظور شناختی	سطح منظور شناختی	√	
۱۸	تفسیر لفظی	تفسیر لفظی	تفسیر لفظی	√	√
۱۹	ربط	ربط	ربط	√	√
۲۰	فهرست واژه‌های غیرمجاز	واژه غیر مجاز	واژه‌های غیرمجاز		
۲۱	تجزیه کننده‌ها	تجزیه کننده	تجزیه کننده‌ها		
۲۲	پایگاه اطلاعاتی واژگانی وُردنیت	پایگاه اطلاعات واژگانی وردنت	وردنت		
۲۳	موجودیت‌ها	موجودیت	موجودیتها		
۲۴	پایگاه اطلاعاتی هستی‌شناسی	هستی شناسی	پایگاه اطلاعاتی هستی شناسی		
۲۵	نظام زبان طبیعی اسمارت	اسمارت (زبان طبیعی اسمارت)	اسمارت		
۲۶	میان زبانی	میان زبانی	میان زبانی	√	√
۲۷	هوش مصنوعی	هوش مصنوعی	هوش مصنوعی	√	√
۲۸	فوق متن	فوق متن	فوق متن	√	√
۲۹	بردارهای اصطلاح	بردار اصطلاح	بردارهای اصطلاح	√	
۳۰	اپراتورهای بولی	اپراتورهای بولی	اپراتورهای بولی	√	√
۳۱	نمایه معکوس	نمایه معکوس	نمایه معکوس	√	√
۳۲	متن آزاد	متن آزاد	متن آزاد	√	√
۳۳	واژه‌های هم نویسه	واژه‌های هم نویسه	واژه‌های هم نویسه	√	√
۳۴	شناسگر	شناسگر	شناسگرها		
۳۵	رتبه‌بند	رتبه بند	رتبه بندها		
۳۶	شمول معنایی	شمول معنایی	شمول معنایی	√	√
۳۷	آلتا ویستا	آلتا ویستا	آلتاویستا	√	
۳۸	همجواری	همجواری	همجواری	√	√
۳۹	جفت جلویی	جفت جلویی	جفت جلویی	√	√
۴۰	جفت عقبی	جفت عقبی	جفت عقبی	√	√

۴۱	شناسگر عبارت اسمی	شناسگر عبارت اسمی	شناسگرها	
۴۲	ریشه‌ساز پورتر	ریشه‌ساز پورتر	ریشه‌ساز پورتر	√
۴۳	دقت	دقت	دقت	√
۴۴	بازیافت	بازیافت	بازیافت	√
۴۵	ماشین خودکار قطعی با حالت منتهی	ماشین خودکار قطعی با حالت منتهی	ماشین خودکار قطعی با حالت نامنتهی	
۴۶	پس‌پردازشگر	پس‌پردازشگر	پس‌پردازشگر	
۴۷	یوروفارسی	یورو فارسی	یورو فارسی	√
۴۸	ماشین خودکار با حالت منتهی	ماشین با حالت منتهی	ماشین خودکار با حالت منتهی	√
۴۹	کوتاه سازی	کوتاه سازی	کوتاه سازی	√

در جدول ۲ در مواردی که در ستون «اصطلاح در متن» بیش از یک واژه یا عبارت دیده می‌شوند این واژگان یا عبارات با شکل‌های گوناگونی در متن استفاده شده‌اند. تلاش شده است 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 آمده است که جایی برای آن‌ها در جدول ۲ از منظر مقایسه با واژه‌نامه وجود نداشته است.

جدول ۳. اصطلاحات دیگر نمایه که در جدول ۲ نیامده است

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

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

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

#### ۴. نتیجه‌گیری و چند پیشنهاد

کتاب پردازش زبان طبیعی و بازیابی اطلاعات اگرچه کتابی است دربارهٔ موضوعی مهم و اساسی برای رشته علم اطلاعات و دانش‌شناسی، ایراداتی دارد که اگر قرار است این کتاب به چاپ‌های بعدی برسد، باید آن‌ها برطرف شوند. اولاً، فصل‌ها بهتر است جابه‌جا 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 «ریشه‌سازی» توضیح داده شده است. فصل هفتم از جمله فصل‌هایی است که باید در ابتدای کتاب بیاید. محتوای برخی از فصل‌ها برای کتاب آموزشی مناسب نیستند. برای نمونه، فصل چهارم صرفاً گزارش تجربه‌ای است و درحالی‌که ترجمهٔ چندان مناسبی هم ندارد با هدف کلی این کتاب هم‌خوان نیست. این خصوصیت درمورد 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 افزون‌بر مواردی که در بالا درمورد ناسلیسی متن ارائه شد، ویرایش ادبی جدی‌ای را می‌طلبد. در صفحات مقدماتی کتاب نیز از هیچ ویرایشی، 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 ویرایش علمی هم‌راه با توضیحات فراوان و ویرایش ادبی جدی نیاز دارد.

## ۵. یادداشت

۱. دربارهٔ مریم ناصری جست‌وجوی اینترنتی هیچ کمکی نکرد. تنها اطلاعی که کسب شد این بود که وی ویراستار فارسی مجلهٔ 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 درحدّ یک پاراگراف اطلاعات اندکی دربارهٔ فعالیت‌ها و آثار علمی خود برای بنده بفرستید. با سپاس فراوان مرتضی کوکی» برای دو گروه بحث مربوط به رشته فرستاد، اما پاسخی دریافت نکرد.

## کتاب‌نامه

*پردازش زبان طبیعی و بازیابی اطلاعات* (۱۳۸۸)، ترجمهٔ جعفر مهاد و مریم ناصری، تهران: چاپار/ شیراز: مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری.

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

