

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه اصفهان

دانشکده فنی و مهندسی

گروه مهندسی کامپیوتر

پایان نامه کارشناسی ارشد رشته مهندسی کامپیوتر گرایش نرم افزار

استخراج روابط مکمل موجود بین کالاها در یک آنتولوژی

استادان راهنما:

دکتر محمدعلی نعمت بخش

دکتر محمدرضا خیام‌باشی

پژوهشگر:

ژیلا نوری

آذر ماه ۱۳۸۸

کلیه حقوق مادی مرتب بر نتایج مطالعات، ابتکارات
و نوآوری‌های ناشی از تحقیق موضوع این پایان نامه
متعلق به دانشگاه اصفهان است.



دانشگاه اصفهان
دانشکده مهندسی
گروه مهندسی کامپیوتر

پایان نامه کارشناسی ارشد رشته مهندسی کامپیوتر گرایش نرم افزار
خانم ژیلا نوری تحت عنوان
استخراج روابط مکمل موجود بین کالاها در یک آنتولوژی

در تاریخ ۸۸/۹/۲۹ توسط هیئت داوران زیر بررسی و با درجه عالی به تصویب نهایی رسید.

۱- استاد راهنمای اول پایان نامه: دکتر محمدعلی نعمت‌بخش با مرتبه علمی دانشیاری امضا

۲- استاد راهنمای دوم پایان نامه: دکتر محمدرضا خیام‌باشی با مرتبه علمی استادیاری امضا

۳- استاد داور داخل گروه: دکتر ناصر نعمت‌بخش با مرتبه علمی استادیاری امضا

۴- استاد داور خارج از گروه: دکتر سید رسول موسوی با مرتبه علمی استادیاری امضا

امضاء مدیر گروه

تقدیم به:

پدر و مادرم مهربان و فداکارم که در تمام مراحل زندگی مرا راهنمایی
و پشتیبانی کرده‌اند.

سپاسگزاری

سپاس خدای یکتا را که هر چه هست از اوست.

سپاس او را که بی نام و یادش نمی‌توان آغاز کرد و نمی‌توان به پایان برد.

سپاس و حمد بیکران او را که به ما طریق عاشقی آموخت.

سپاس ترا ای یگانه عالم، ای حقیقت مطلق، ای منتهای تحقیق، ای همه هستی در یک کلام

بر خود فرض میدانم تا از تلاش تمامی راهنمایان خود در این پروژه به ویژه جناب آقای دکتر محمدعلی نعمت‌بخش و دکتر محمدرضا خیام‌باشی تشکر کنم و از پروردگار یکتا توفیق روز افزون ایشان را مسئلت دارم.

از خانواده مهربان و فداکارم که با تشویق نمودن و تحمل دشواری‌ها، یاریگر و پشتیبان اینجانب در این راه بوده‌اند، سپاسگزاری کرده و از خداوند بزرگ سلامت ایشان را طلب می‌کنم.

چکیده

در سال‌های اخیر، تحقیقات بسیاری جهت پیدا کردن روابط معنایی بین موجودیت‌ها به منظور بالا بردن سطح تجاری ارگان‌ها و شرکت‌ها انجام شده است. یکی از این روابط، رابطه مکمل بین کالاهاست. این رابطه، کالاهایی را مشخص می‌کند که در جهت برآورده ساختن یک نیاز با یکدیگر مصرف می‌شوند. این رابطه کاربردهای فراوانی در زمینه تجارت الکترونیکی و بازاریابی دارد. در این رساله، اقدام جدیدی جهت کشف این روابط انجام گرفته است. بدین صورت که یک روش خودکار و نظارت نشده جهت استخراج روابط مکمل بین کالاهای موجود در یک کاتالوگ محصولات در قالب یک روش یادگیری آنتولوژی با استفاده از اسناد موجود در وب به عنوان مجموعه نوشته‌های یادگیری طراحی و پیاده‌سازی شده است. در عصر حاضر، با رشد سریع جامعه اطلاعاتی، وب تبدیل به مخزن اطلاعاتی با ارزشی در همه حوزه‌های دانشی ممکن جهت انجام وظایف کسب دانش و بازاریابی اطلاعات شده است. روش یادگیری آنتولوژی ارائه شده از خصوصیات وب به عنوان مجموعه نوشته‌های یادگیری بهره‌برداری می‌کند و بدین منظور روش‌های جدیدی جهت کسب دانش از مجموعه نوشته‌های وب، مورد استفاده قرار گرفته‌اند. رویه ارزیابی خودکاری نیز جهت ارزیابی روابط استخراج شده با استفاده از مخزن اطلاعاتی WordNet، طراحی شده است که روابط استخراج شده با استفاده از روش یادگیری آنتولوژی پیشنهادی به وسیله این رویه، مورد ارزیابی قرار گرفته است.

از آن جایی که کالاهای مکمل کاربردهای فراوانی در بازاریابی و تجارت الکترونیک دارند و جهت نشان دادن کاربرد روابط استخراج شده، در نهایت یکی از کاربردهای این کالاها در زمینه تبلیغات زمینه‌ای انتخاب شده است. تبلیغات زمینه‌ای، یکی از انواع تبلیغات اینترنتی است که به صورت خودکار براساس محتوای صفحات وبی که کاربر در حال مرور کردن آن است، تبلیغات مرتبط را در این صفحات قرار می‌دهد. با استفاده از روابط استخراج شده از مرحله قبل، الگوریتم جدیدی جهت تطابق تبلیغات و صفحات در تبلیغات زمینه‌ای طراحی و پیاده‌سازی شده و سپس کارایی روابط استخراج شده در این تبلیغات، مورد ارزیابی قرار گرفته است.

کلید واژه‌ها: روابط معنایی، کالای مکمل، یادگیری آنتولوژی، وب کاوی، تبلیغات زمینه‌ای

فهرست مطالب

صفحه

عنوان

فصل اول: مقدمه

- ۱-۱ معرفی آنتولوژی ۴
- ۱-۱-۱ اساس کار آنتولوژی ۵
- ۲-۱ منبع یادگیری جدید: وب ۸
- ۳-۱ اهداف و کاربردهای تحقیق ۱۰
- ۴-۱ ساختار پایان نامه ۱۲

فصل دوم: پیشینه و تاریخچه موضوع تحقیق

- ۱-۲ استخراج اطلاعات ۱۴
- ۲-۲ کسب دانش از متون ۱۸
- ۳-۲ خلاصه و رابطه با طرح پیشنهادی ۲۴

فصل سوم: منابع و روش‌های استفاده شده جهت یادگیری

- ۱-۳ وب به عنوان یک مجموعه نوشته یادگیری ۲۹
- ۲-۳ کاتالوگ محصولات ۲۹
- ۳-۳ روش تحلیل خفیف و ابزارهای پردازش زبان طبیعی ۳۱
- ۴-۳ تحلیل آماری ۳۲
- ۵-۳ موتورهای جستجوی وب ۳۴
- ۱-۵-۳ طبقه‌بندی موتورهای جستجوی وب ۳۵
- ۲-۵-۳ موتورهای جستجوی وب به عنوان ابزارهای یادگیری ۳۹
- ۳-۵-۳ مقایسه موتورهای جستجوی مبتنی بر کلمه کلیدی ۴۱
- ۶-۳ خلاصه فصل و نتیجه‌گیری ۴۴

فصل چهارم: طرح پیشنهادی جهت استخراج روابط مکمل بین کالاها

- ۱-۴ کشف و انتخاب الگوها ۴۸
- ۲-۴ بازیابی و انتخاب مفاهیم مکمل ۵۲
- ۱-۲-۴ اندازه مناسب مجموعه نوشته‌ها ۵۴

۳-۴ محاسبه درجه مکملی بین دو کالا ۵۷

۴-۴ خلاصه فصل و نتیجه‌گیری ۶۱

فصل پنجم: ارزیابی روابط مکمل استخراج شده

۱-۵ معیارهای کیفیت و ضوابط ارزیابی عمومی ۶۴

۲-۵ معرفی معیار ارزیابی طراحی شده ۶۶

۳-۵ ارزیابی روابط مکمل استخراج شده ۷۰

۴-۵ خلاصه فصل و نتیجه‌گیری ۷۳

فصل ششم: کاربرد رابطه مکمل معنایی در تبلیغات زمینه‌ای

۱-۶ تبلیغات اینترنتی ۷۶

۱-۱-۶ تاثیر تبلیغات بر رفتار خرید مصرف کننده ۷۶

۲-۱-۶ انواع تبلیغات اینترنتی ۷۷

۲-۶ انواع تبلیغات متنی ۸۰

۳-۶ اصول و ساختار تبلیغات زمینه‌ای ۸۲

۴-۶ استفاده از روابط معنایی جهت بهبود کارایی تبلیغات زمینه‌ای ۸۵

۱-۴-۶ مدل سازی تبلیغات زمینه‌ای ۸۵

۲-۴-۶ ارائه الگوریتمی جهت تطابق معنایی تبلیغات و صفحات ۸۶

۵-۶ ارزیابی کارایی روش ارائه شده ۸۸

۶-۶ کاربردهای دیگر کالاهای مکمل ۸۹

۷-۶ خلاصه فصل و نتیجه‌گیری ۹۱

فصل هفتم: نتیجه‌گیری و پیشنهاداتی جهت مطالعات آتی

۱-۷ شرح روش و اهمیت نتایج ۹۳

۲-۷ مزایا و کاربردهای روش پیشنهادی ۹۴

۳-۷ راه کارهایی جهت مطالعات آتی ۹۵

پیوست الف: مروری بر ابزار پیاده‌سازی GATE

الف-۱ زمینه‌های کاربرد GATE ۹۸

الف-۲ مرور کلی بر ساختار GATE ۹۹

- الف-۳ مولفه‌های تعیین شده در محیط GATE ۱۰۱
- الف-۴ امکانات استفاده شده از ابزار GATE در پیاده‌سازی تحقیق ۱۰۲
- واژه‌نامه انگلیسی به فارسی ۱۰۳
- منابع و مأخذ ۱۰۷

فهرست شکل‌ها

صفحه	عنوان
۶	شکل ۱-۱ نمونه‌ای از دسته‌بندی آنتولوژی‌ها
۳۰	شکل ۱-۳ بخشی از کاتالوگ محصولات آمازون
۳۶	شکل ۲-۳ گروه‌های تعریف شده به صورت دستی ارائه شده به وسیله یاهو برای شیر خوراکی
۳۸	شکل ۳-۳ خوشه‌هایی از منابع وب مطرح شده به وسیله clusty برای coffee
۳۸	شکل ۴-۳ خوشه‌هایی از منابع وب مطرح شده به وسیله carrot برای hamburger
۴۰	شکل ۵-۳ آمارهایی درباره واژه‌های پرس‌وجو شده در وب برگردانده شده توسط موتور گوگل
۴۷	شکل ۱-۴ بخشی از کاتالوگ محصولات آمازون در قالب یک آنتولوژی حوزه‌ای
۵۱	شکل ۲-۴ خلاصه‌ای از مراحل الگوریتم انتخاب افعال
	شکل ۳-۴ میانگین تعداد عبارات اسمی اضافه شده در هر تکرار فرآیند یادگیری در برابر تعداد منابع وب برگردانده شده توسط گوگل
۵۵	
۶۰	شکل ۴-۴ خلاصه‌ای از مراحل بازیابی و انتخاب کالاهای مکمل
۷۲	شکل ۱-۵ خلاصه‌ای از معیارهای ارزیابی روابط مکمل برای سه کلاس محصولاتی انتخاب شده
۸۳	شکل ۱-۶ موجودیت‌های شرکت کننده در تبلیغات زمینه‌ای و تعامل بین آن‌ها
۸۸	شکل ۲-۶ خلاصه‌ای از مراحل کار الگوریتم تطابق تبلیغات و صفحات
۱۰۱	شکل الف-۱ نمونه‌ای از صفحه اجرای منابع GATE

فهرست جدول‌ها

عنوان	صفحه
جدول ۱-۲ مقایسه مزایا و معایب روش‌های استخراج اطلاعات سنتی و وقتی	۱۵
جدول ۲-۲ خلاصه‌ای از روش‌های اصلی کسب دانش جهت استخراج روابط از متون	۲۱
جدول ۳-۲ خلاصه‌ای از ابزارهای اصلی کسب دانش از متون	۲۲
جدول ۱-۳ خلاصه‌ای از چند موتور جستجوی مبتنی بر خوشه	۳۷
جدول ۲-۳ تعداد نتایج تخمینی به دست آمده به وسیله موتورهای جستجوی وب مبتنی بر کلمه کلیدی برای حوزه‌های عمومی	۴۲
جدول ۳-۳ تعداد نتایج تخمینی به دست آمده به وسیله موتورهای جستجوی وب مبتنی بر کلمه کلیدی برای پرس‌وجوهای خاص	۴۳
جدول ۴-۳ خلاصه‌ای از خصوصیات اصلی موتورهای جستجوی وب مبتنی بر کلمه کلیدی	۴۳
جدول ۱-۴ الگوهای لغوی کشف شده جهت استخراج مفاهیم مکمل در هر کلاس محصولاتی	۵۲
جدول ۲-۴ تعریف عبارات اسمی	۵۲
جدول ۳-۴ مثالی از فرآیند بازیابی کالاهای مکمل	۵۹
جدول ۴-۴ نمونه‌هایی از کالاهای مکمل استخراج شده همراه با درجه مکملی آن‌ها	۵۹
جدول ۱-۵ دسته‌بندی معیارهای تشابه و وابستگی معنایی و مزایا و معایب این روش‌ها	۶۸
جدول ۲-۵ نتایج حاصل از ارزیابی کلاس محصولاتی پوشاک، کفش و جواهرات	۷۱
جدول ۳-۵ نتایج حاصل از ارزیابی کلاس محصولاتی نوشیدنی‌ها	۷۱
جدول ۴-۵ نتایج حاصل از ارزیابی کلاس محصولاتی خوراکی‌ها	۷۲
جدول ۱-۶ نتایج حاصل از ارزیابی الگوریتم ارائه شده جهت تطابق تبلیغات و صفحات	۸۹

فهرست کوتاه نوشت‌ها

XML	Extensible Markup Language
HTML	Hypertext Markup Language
SHOE	Simple HTML Ontology Extensions
DHTML	Dynamic HTML
RTF	Rich Text Format
API	Application Programming Interface
IP	Internet Protocol
NP	Noun Phrase
DT	Determiner
PREP	Preposition
JJ	Adjective
VBG	Verb Gerund
NN	Noun
NNS	Pluran Noun
SynSets	Synonym Sets
IC	Information Content
LCS	Least Common Subsumer
POS	Ports of Speech
Ad	Advertise
KEA	Keyphrase Extraction Algorith
GATE	General Architecture for Text Engineering
JAR	Java Archive
SGML	Standard Generilized Markup Language

فصل اول

مقدمه

امروزه، بسیاری از شرکت‌ها و ارگان‌های تجاری به منظور بالا بردن سطح تجاری خود سرمایه‌گذاری وسیعی در بخش تجارت الکترونیکی انجام می‌دهند. این شرکت‌ها حجم وسیعی از داده‌ها را در پایگاه داده‌های خود نگهداری می‌کنند؛ اما این داده‌ها زمانی مفید هستند که روابط^۱ موجود بین آن‌ها کشف شوند. بدون کشف این روابط، فقط داده‌های خامی در اختیار این شرکت‌ها است که طبیعتاً قدرت استفاده از این اطلاعات را جهت بهبود استراتژی فروش، سرمایه‌گذاری، و مدیریت سیاست بازاریابی نخواهند داشت [۱].

از گذشته، روابط موجود بین کالاها در تحقیقات علمی تجاری و بازاریابی به منظور بهبود استراتژی تجارت، مورد استفاده قرار می‌گرفتند تا بتوانند کارایی سازمان‌ها، سیاست‌های مدیریت، استراتژی فروش، سرمایه‌گذاری و بازاریابی را بهبود بخشند و همچنین رشد قابل قبول دراز مدتی را در تجارت ایجاد کنند. در ادامه تعدادی از انواع روابط موجود بین کالاها معرفی شده‌اند.

یکی از روابط موجود بین کالاها که در تحقیقات مربوط به داده کاوی معرفی شده است، رابطه خرید^۲ است. در یک مجموعه از تراکنش‌ها که هر تراکنش شامل مجموعه‌ای از اقلام (محصولات یا سرویس‌ها) است، این رابطه اقلامی را مشخص می‌کند که معمولاً در تراکنش‌های خرید با یکدیگر همراه هستند و قوانین انجمنی^۳ را بیان می‌-

1 Associations
2 Purchasing Association
3 Association Rules

کنند. یک قانون انجمنی عبارتی به فرم $x \Rightarrow y$ است که در آن x و y ، مجموعه‌ای از اقلام هستند. معنی شهودی این قانون این است که تراکنش‌هایی در پایگاه داده که شامل اقلام موجود در x هستند، مایل به دارا بودن اقلام موجود در y نیز هستند. به عنوان مثال ممکن است قانونی به این صورت بیان شود: ۹۸٪ از مشتریانی که لاستیک و لوازم یدکی خودرو را خریداری می‌کنند، برخی از سرویس‌های خودرو را نیز خریداری می‌کنند [۲]. در واقع این رابطه رفتار خرید مشتریان را تحت بررسی قرار می‌دهد اما لزوماً اقلامی که با یکدیگر مصرف می‌شوند را مشخص نخواهد کرد. معایب اصلی این رابطه عبارتند از دقت و کارایی پایین، و رشد نمایی روابط با افزایش تعداد اقلام [۱].

رابطه دیگری که بین کالاها وجود دارد، رابطه تشابه^۱ است. این رابطه درجه شباهت بین کالاها را نشان می‌دهد. کالاهای مشابه، کالاهایی هستند که از نظر کاربرد و خصوصیات به یکدیگر شبیه هستند یا در واقع وابستگی معنایی بیشتری با یکدیگر دارند. به عنوان مثال ماشین و دوچرخه دارای شباهت‌های بیشتری نسبت به ماشین و بنزین هستند در نتیجه وابستگی تشابه معنایی بین ماشین و دوچرخه بیشتر بوده و از درجه تشابه بالاتری برخوردار خواهند بود. تا کنون روش‌ها و معیارهای بسیاری جهت محاسبه درجه تشابه بین اقلام ارائه شده است [۳].

نوع دیگری از روابط موجود بین کالاها، رابطه جایگزینی^۲ است که در سال‌های اخیر مورد توجه بسیاری قرار گرفته است. این رابطه، کالاهایی را مشخص می‌کند که می‌توانند به جای یکدیگر مورد استفاده قرار گیرند. به عنوان مثال چای و قهوه دو کالای جایگزین هستند که در نبود یکی می‌توان از دیگری استفاده کرد. معمولاً مصرف‌کنندگان در ساختار اولویت‌هایشان به دو کالای جایگزین میزان سودمندی یکسانی می‌دهد، بنابراین لزومی ندارد که این دو کالا را باهم و در یک زمان خریداری کنند [۴]. در اصل می‌توان گفت کالاهای جایگزین، کالاهایی هستند که درجه شباهت بالایی دارند اما درجه رابطه خرید بین آن‌ها به طور تقریبی صفر است.

رابطه دیگری بین کالاها وجود دارد که متفاوت از روابط مطرح شده قبلی است، به عنوان مثال رابطه بین ماشین و بنزین را در نظر بگیرید به نظر می‌آید که رابطه نزدیکی با هم دارند، اما شبیه یا جایگزین هم نیستند ولی می‌توان گفت برای استفاده از ماشین نیاز به بنزین است. بنابراین رابطه دیگری بین کالاها وجود دارد که در آن یک کالا

1 Similarity Association

2 Substitution Association

هنگام مصرف، نیاز کالای دیگر را برطرف می‌کند. این رابطه در علم اقتصاد به عنوان رابطه مکملی^۱ تعریف شده است. با کمک رابطه مکملی بین کالاها می‌توان به تعریف جدیدی به نام کالای مکمل^۲ رسید. کالای مکمل، به معنی کالایی است که با کالای دیگر مصرف می‌شود و البته لزوماً به این معنی نیست که این کالاها باهم خریداری می‌شوند. نمونه‌ای دیگر از کالای مکمل می‌تواند همبرگر و نان همبرگر باشد که نان همبرگر، یکی از نیازهای همبرگر را برطرف می‌کند و معمولاً با همبرگر مصرف می‌شود [۵، ۶].

اگر قیمت یک کالا کاهش یابد و مصرف‌کنندگان مقدار بیشتری از این کالا را خریداری کنند، معمولاً آن‌ها مقدار بیشتری از کالای مکمل آن کالا را نیز، صرف نظر از کاهش یا عدم کاهش قیمت کالای مکمل، خریداری خواهند کرد. به طور مشابه اگر قیمت یک کالا افزایش یافته و میزان تقاضا برای آن کالا کاهش یابد، میزان تقاضای کالای مکمل آن کالا نیز کاهش خواهد یافت [۷]. در بسیاری از برنامه‌های کاربردی تجارت الکترونیکی این نوع رابطه مفیدتر از روابط دیگر است، به عنوان مثال، در سیستم‌های پیشنهاددهنده، پیشنهاد دادن یک ماشین مشابه یا دوچرخه به کسی که یک ماشین مرسدس خریده است زیاد مفید نیست، اما پیشنهاد یک ضبط صوت یا سیستم دزدگیر به عنوان کالای مکمل ماشین، مفیدتر است.

کالاهای مکمل کاربردهای فراوانی در زمینه تبلیغات، بازاریابی، اقتصاد و فروش، سیاست و موضوع‌های اجتماعی دارند. به عنوان مثال در بازاریابی، کالاهای مکمل قدرت بازاریابی بیشتری به شرکت‌ها می‌دهند چنانچه اگر فروشنده‌ای بخواهد میزان فروش کالایی را بالا ببرد، کالاهای مکمل آن کالا را تبلیغ می‌کند.

اگر چه کالاهای مکمل از اهمیت بالایی برخوردار هستند، با توجه به تحقیقات انجام شده تاکنون تحقیقی به منظور یافتن کالاهای مکمل به طور خودکار انجام نگرفته است. در یکی از تحقیقات انجام شده، از کالاهای مکمل که به صورت دستی تعیین شده‌اند، در یک سیستم پیشنهاد دهنده استفاده شده است که نتایج ارزیابی در این تحقیق نشانگر بهبود کارایی و دقت بوده است [۸]. نویسندگان اثبات کرده‌اند که روششان بسیار کاراتر از قوانین انجمنی بوده و مشکلات مقیاس‌پذیری و آغاز سرد را حل کرده است. متأسفانه مشکل اساسی روش مطرح شده این است که کالاهای مکمل به صورت دستی مشخص شده‌اند.

1 Complementary Association

2 Complementary Product

در این رساله، روش جدیدی جهت پیدا کردن روابط مکمل موجود بین کالاهای^۱ و نظارت نشده^۲ با استفاده از روشی مبتنی بر یادگیری آنتولوژی^۳ در قالب یک رابطه معنایی طبقه‌بندی نشده^۴ بین مفاهیم ارائه شده است. در ادامه مفاهیم آنتولوژی و یادگیری آنتولوژی به طور مختصر معرفی شده‌اند.

۱-۱ معرفی آنتولوژی

در انتهای قرن بیستم و ابتدای قرن بیست و یکم، آنتولوژی‌ها به عنوان زمینه تحقیقاتی مهمی وارد علوم کامپیوتری شدند. آنتولوژی کلمه‌ای یونانی به معنی هستی‌شناسی است که در مورد نظریه وجود بحث می‌کند. نظریه‌ای در فلسفه که می‌تواند به سئوالاتی از قبیل «چگونگی وجود»، «ویژگی‌های مشترک در همه موجودات»، «ویژگی‌هایی که می‌توانند وجود را توصیف کنند» پاسخ دهد.

در عصر حاضر، آنتولوژی‌ها به منظور اشتراک و استفاده مجدد دانش در میان حوزه‌ها^۵ و وظایف^۶ ایجاد می‌شوند. در حال حاضر، آن‌ها به طور گسترده در مهندسی دانش، هوش مصنوعی و علوم کامپیوتر در کاربردهایی مربوط به مدیریت دانش، پردازش زبان طبیعی، تجارت الکترونیکی، بازیابی اطلاعات، یکپارچه‌سازی و طراحی پایگاه داده، وب معنایی، آموزش و غیره مورد استفاده قرار می‌گیرند. یکی از اهداف آن‌ها کاهش (یا حذف) اختلال مفهومی و اصطلاحی بین اعضای یک اجتماع مجازی از کاربرانی (افراد یا برنامه‌های کامپیوتری) است که نیاز به اشتراک گذاشتن اسناد الکترونیکی و گونه‌های اطلاعاتی متفاوتی دارند. این امر به واسطه شناسایی و تعریف مجموعه‌ای از مفاهیم مربوط که یک حوزه کاربردی معین را توصیف می‌کنند، حاصل می‌شود. در نتیجه برخی از دلایل توصیف آنتولوژی‌ها عبارتند از [۹]:

- تصریح کردن مفروضات حوزه‌ای و ساده‌تر کردن جهت تغییر و ادراک.
- تفکیک کردن دانش حوزه‌ای از دانش عملیاتی.
- تشکیل یک مرجع عمومی برای کاربردها.
- به اشتراک گذاشتن یک ادراک سازگار از معانی اطلاعات.

1 Automatic
 2 Unsupervised
 3 Ontology Learning
 4 Non-Taxonomic Semantic Relationship
 5 Domains
 6 Tasks

تعاریف مختلفی برای آنتولوژی وجود دارد؛ استادر و همکارانش، آنتولوژی را به عنوان یک مشخصه صریح قراردادی از یک تصور مشترک تعریف کرده اند [۱۰]. در این تعریف، تصور اشاره به یک مدل انتزاعی از پدیده‌هایی در دنیا دارد که مفاهیم مربوط به آن پدیده شناسایی شده‌اند. صریح بدین معنی است که نوع مفاهیم استفاده شده و قیود استفاده آن‌ها باید به صراحت تعریف شوند. قراردادی اشاره به این اصل دارد که آنتولوژی باید قابل خواندن توسط ماشین باشد و مشترک حاکی از نظریه‌ای است که آنتولوژی باید دانش رضایتی^۱ را به تصرف در آورد، بدین معنی که نه تنها توسط افرادی بلکه توسط گروهی پذیرفته شده باشد.

تعریف دیگری توسط نجس و همکارانش با توجه به قالب یک آنتولوژی ارائه شده است؛ یک آنتولوژی، واژه‌ها و روابط اساسی را تعریف می‌کند که شامل واژه نامه‌ای از حوزه یک موضوع و همچنین قوانینی به منظور ترکیب واژه‌ها و روابط جهت تعریف توسعه‌ها در واژه نامه است [۱۱].

۱-۱-۱ اساس کار آنتولوژی

اگرچه روش‌های نمایش دانش مختلفی برای تعریف آنتولوژی‌ها وجود دارد، اما همه آن‌ها مجموعه کمینه مولفه-های زیر را به اشتراک می‌گذارند:

- کلاس‌ها^۲: که مفاهیم^۳ را نشان می‌دهند. کلاس‌ها در یک آنتولوژی، معمولاً به صورت طبقه‌بندی شده^۴ به واسطه مکانیزم‌های وراثتی که می‌توانند اعمال شوند، سازماندهی می‌شوند.
- روابط^۵: نوعی وابستگی بین مفاهیم یک حوزه را نشان می‌دهند. آنتولوژی‌ها معمولاً شامل روابط دودویی هستند. آرگومان اول نشان دهنده دامنه رابطه و آرگومان دوم نشان دهنده برد^۶ رابطه است. روابط دودویی گاهی جهت نمایش صفات مفاهیم نیز استفاده می‌شوند ولی تفاوت آن‌ها با روابط در برد آن‌هاست که یک نوع داده‌ای مانند رشته‌ای، عددی و غیره است، در حالی که برد رابطه یک مفهوم است.
- نمونه‌ها^۷: جهت نمایش عناصر و افراد در یک آنتولوژی استفاده می‌شوند.

1 Consensual Knowledge

2 Classes

3 Concepts

4 Taxonomic

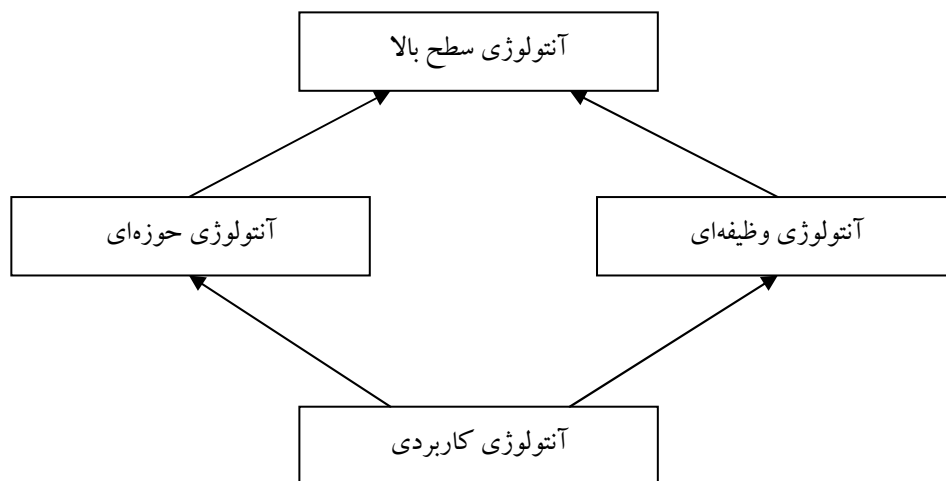
5 Relations

6 Range

7 Instances

در حال حاضر دسته‌بندی‌های مختلفی برای آنتولوژی‌ها بر حسب یک جنبه خاص (مانند موضوع یا نوع ساختار) در دسترس است. دسته بندی جالبی که توسط گوارینو مطرح شده است، انواع آنتولوژی‌ها را با توجه به سطح وابستگی آن‌ها به یک وظیفه خاص یا یک نقطه نظر چنانچه در شکل ۱-۱ نشان داده شده است، دسته بندی کرده است [۱۲]:

- آنتولوژی‌های سطح بالا^۱: مفاهیم بسیار عمومی مانند فضا، زمان و رخداد را توصیف می‌کنند که مستقل از یک حوزه خاص هستند. داشتن آنتولوژی‌های سطح بالای متحد برای اجتماع بزرگی از کاربران معقول به نظر می‌رسد. مثال‌هایی از این آنتولوژی‌ها عبارتند از: Cyc's، Sow's و SUO.
- آنتولوژی‌های حوزه‌ای^۲: واژه‌نامه‌ای مربوط به یک حوزه عمومی را به وسیله اختصاصی کردن مفاهیم معرفی شده در آنتولوژی سطح بالا، توصیف می‌کنند.
- آنتولوژی‌های وظیفه‌ای^۳: واژه‌نامه‌ای مربوط به یک وظیفه یا فعالیت عمومی را به وسیله اختصاصی کردن آنتولوژی‌های سطح بالا، توصیف می‌کنند.
- آنتولوژی‌های کاربردی^۴: این دسته از آنتولوژی‌ها، خاص‌ترین نوع آنتولوژی‌ها هستند. مفاهیم در آنتولوژی‌های کاربردی معمولاً متناظر با نقش‌های ایفا شده به وسیله موجودیت‌های حوزه هستند.



شکل ۱-۱ نمونه‌ای از دسته‌بندی آنتولوژی‌ها [۱۲]

1 Top-level Ontologies
 2 Domain Ontologies
 3 Task Ontologies
 4 Application Ontologies

به مجموعه فعالیت‌های مرتبط با فرآیند توسعه آنتولوژی، چرخه حیات آنتولوژی، اصول و روش‌های ساخت آنتولوژی‌ها، و ابزار مناسب و زبان‌هایی که آن‌ها را پشتیبانی می‌کنند، مهندسی آنتولوژی گویند [۹]. با توجه به اسلوب‌ها، طرح‌های متفاوتی برای توسعه آنتولوژی‌ها به صورت دستی ارائه شده است.

با توجه به دسته‌بندی ارائه شده از آنتولوژی‌ها در بالا، فلاسفه آنتولوژی‌دان و منطق‌دانان هوش مصنوعی، معمولاً درگیر وظیفه تعریف انواع و ساختارهای اصلی تغییر ناپذیر مفاهیمی (اشیا، ویژگی‌ها، روابط و قضایا) هستند که در هر حوزه ممکن قابل کاربرد اند. این اصول اصلی معمولاً در آنتولوژی‌های سطح بالا قرار می‌گیرند.

از طرف دیگر، آنتولوژی‌های کاربردی دارای زمینه و قابلیت استفاده مجدد بسیار محدودی هستند، به دلیل این که آن‌ها وابسته به یک حوزه خاص بوده و نیازمند یک کاربرد ویژه هستند. این آنتولوژی‌ها معمولاً بنا به هدف ویژه‌ای توسط طراحان برنامه‌های کاربردی توسعه می‌یابند.

در یک نقطه میانه، آنتولوژی‌های وظیفه‌ای و حوزه‌ای پیچیده‌ترین نوع آنتولوژی‌ها از لحاظ توسعه هستند؛ به این دلیل که از یک طرف آن‌ها به اندازه‌ای عمومی هستند که نیازمند به دست آوردن اجماع^۱، بین اجتماع بزرگی از کاربران هستند و از طرف دیگر به اندازه‌ای منسجم^۲ هستند که می‌توانند تنوع بزرگی با حوزه‌های دانش بسیار متفاوت و پویا با میلیون‌ها مفهوم ممکن جهت مدل کردن را ارائه دهند.

ابتکار جدیدی مانند وب معنایی به شدت به آنتولوژی‌های حوزه‌ای متکی است. تلاش وب معنایی به دست آوردن وب تفسیر شده به صورت معنایی است که در آن موتورهای جستجو بتوانند اطلاعات موجود در منابع وب را از نقطه نظر معنایی پردازش کنند و کیفیت اطلاعات ارائه شده به کاربران را به شدت افزایش دهند [۱۳]. این روش نیازمند اجماعی سراسری در تعریف ساختارهای معنایی مناسب (آنتولوژی‌های حوزه‌ای) جهت نمایش هر حوزه ممکن از دانش است. در نتیجه، به طور گسترده‌ای توافق شده است که توده بحرانی از آنتولوژی‌ها جهت نمایش معنا در وب معنایی مورد نیاز است.

ساخت آنتولوژی‌های حوزه‌ای به وسیله طراحان حوزه و مهندسين دانش انجام می‌گیرد که معمولاً تحت تاثیر مشکلاتی از قبیل اندازه بالقوه، پیچیدگی و پویایی یک حوزه خاص قرار می‌گیرند. در نتیجه ساخت یک آنتولوژی حوزه‌ای کامل دارای مشکلاتی است که پروژه‌های بسیار کمی توانسته‌اند بر آن‌ها غلبه کنند. بنابراین،

1 Consensus

2 Concrete