Downloaded from jor.iranaict.ir at 0:57 +0330 on Monday November 9th 2026

سال چهارم، شماره‌های ۱۱ و ۱۲، بهار و تابستان ۱۳۹۱
صص:۴۷-۵۴

فصلنامه علمی- پژوهشی
فناوری اطلاعات و ارتباطات ایران

ارائه مدلی برای استخراج اطلاعات از مستندات متنی، مبتنی بر متن‌کاوی در حوزه یادگیری الکترونیکی

احمد آقا کاردان* مینا کیهانی نژاد**[1]

* استادیار، دانشگاه امیرکبیر (پلی تکنیک تهران)، آزمایشگاه پیشرفته فناوری‌های یادگیری الکترونیکی

** دانشجوی کارشناس ارشد، دانشگاه امیرکبیر (پلی تکنیک تهران)، آزمایشگاه پیشرفته فناوری‌های یادگیری الکترونیکی

تاریخ دریافت: ۱۳۹۱/۰۲/۲۰ تاریخ پذیرش: ۱۳۹۱/۱۱/۲۳

چکیده

هنگامی‌که شبکه‌های کامپیوتری ستون اصلی علم و اقتصاد شد، حجم زیادی از مستندات در دسترس قرار گرفتند. به همین منظور، برای استخراج اطلاعات مفید از روش‌های متن‌کاوی استفاده می‌شود. متن‌کاوی یک حوزه پژوهشی مهم در کشف اطلاعات ناشناخته، فرضیات، و حقایق جدید به‌وسیله استخراج اطلاعات از اسناد مختلف است. همچنین متن-کاوی آشکار کردن اطلاعات پنهان با استفاده از روشی است که در یک طرف توانایی مقابله با تعداد زیادی کلمات و ساختارهایی در زبان طبیعی را نشان می‌دهد و از طرف دیگر اجازه مدیریت ابهام و شک را می‌دهد. علاوه بر آن، متن‌کاوی به عنوان داده‌کاوی متن بیان می‌شود که معادل با تجزیه و تحلیل متون است و به فرایند استخراج اطلاعات از متن می‌پردازد و اطلاعات با کیفیت بالا را از میان الگوها و فرایندها استخراج می‌کند. همچنین به عنوان داده‌کاوی متن یا کشف دانش از پایگاه داده‌های متنی شناخته می‌شود و به فرایند استخراج الگوها یا دانش از اسناد متنی بیان می‌شود. روش تحقیق در این کار بدین صورت است که ابتدا به بررسی پژوهش‌های انجام شده در حوزه متن‌کاوی با تأکید بر روش‌ها و کاربردهای آن در آموزش الکترونیکی پرداخته شد. در طی این مطالعات، پژوهش‌های مرتبط در حوزه آموزش الکترونیکی طبقه‌بندی گردیدند. پس از طبقه‌بندی پژوهش‌ها، مسائل و راهکارهای مرتبط با مسائل مطرح شده در آن کارها، استخراج شدند. در همین راستا، در این مقاله ابتدا به تعریف متن‌کاوی پرداخته می‌شود. سپس فرایند متن‌کاوی و حوزه‌های کاربرد متن‌کاوی در آموزش الکترونیکی مورد بررسی قرار می‌گیرند. در ادامه روش‌های متن‌کاوی معرفی شده و تک تک این روش‌ها در حوزه آموزش الکترونیکی مطرح می‌گردد. در انتها ضمن استنتاج نکات مهم مطالعات انجام شده، مدلی جهت استخراج اطلاعات برای بهره‌برداری از روش‌های متن‌کاوی در یادگیری الکترونیکی پیشنهاد می‌شود.

کلید واژگان: داده‌کاوی، متن‌کاوی، آموزش الکترونیکی، یادگیری الکترونیکی، سیستم مدیریت یادگیری

۱- مقدمه

داده‌کاوی یا کشف دانش در پایگاه داده (KDD[1])، استخراج خودکار الگوهای جالب و مفهومی از مجموعه داده بزرگ است. بعضی وظایف و روش‌های مفید داده‌کاوی، آمار، مجازی‌سازی، خوشه‌بندی، طبقه‌بندی و قوانین انجمنی هستند[۱]. داده‌کاوی را می‌توان به عنوان فرایند استخراج اطلاعات مفید از داده تعریف کرد[۲]. متن‌کاوی، دگرگونی روی زمینه‌ای است که داده‌کاوی نامیده می‌شود و سعی بر یافتن الگوهای جالب از پایگاه داده‌های بزرگ دارد. متن‌کاوی همچنین به عنوان تحلیل متن هوشمند، شناخته می‌شود. متن داده‌کاوی یا کشف دانش در متن (KDT[2])، عموماً به فرایند استخراج اطلاعات و دانش جالب از متون ساخت‌نیافته اشاره دارد[۳].

متن‌کاوی یا کشف دانش از متن، برای اولین بار توسط Feldman و همکاران [FD95] به آن توجه شد. روش‌های بازیابی و استخراج اطلاعات در متن‌کاوی از پردازش زبان طبیعی (NLP[3]) استفاده می‌کنند و آن‌ها را با

[1] Knowledge Discovery in Database

[2] Knowledge Discovery in Text

[3] Natural Language Processing

[1] نویسنده عهده‌دار مکاتبات (mina_keyhaninejad@aut.ac.ir)

Downloaded from jor.iranaict.ir at 0:57 +0330 on Monday November 9th 2026

الگوریتم‌ها و روش‌های KDD، داده‌کاوی، ماشین یادگیری و آمار مرتبط می‌کنند. بنابراین، یک رویه مشابه با فرآیند KDD انتخاب می‌شود. متن کاوی می‌تواند همچنین مشابه داده‌کاوی به عنوان کاربرد الگوریتم‌ها و روش‌ها در زمینه یادگیری ماشین و آمار به متن با هدف یافتن الگوهای مفید تعریف شود. به همین منظور لازم است نتیجتاً متن پیش پردازش شود. بسیاری نویسنده‌ها روش‌های استخراج اطلاعات و پردازش زبان طبیعی را به منظور استخراج داده از متن استفاده می‌کنند. متن‌کاوی، کشف به‌وسیله اطلاعات ناشناخته قبلی و استخراج خودکار اطلاعات از منابع نوشته شده مختلف است[٤].

متن‌کاوی شبیه داده‌کاوی است، به جز اینکه ابزارهای داده‌کاوی برای مدیریت داده ساخت یافته پایگاه داده‌ها طراحی می‌شود، اما متن‌کاوی می‌تواند با داده نیمه ساخت یافته یا ساخت نیافته مانند پست الکترونیکی، اسناد متون کامل و فایل‌های HTML، سروکار داشته باشد. در نتیجه، متن-کاوی یک راه حل بسیار بهتر برای سازمان‌ها است. با این وجود، بیشتر تحقیقات و تلاش‌های گسترده روی کوشش‌های داده‌کاوی که داده ساخت یافته استفاده می‌کند، مرکزیت می‌یابد[٣].

٢- فرایند متن‌کاوی

در این قسمت فرایند متن‌کاوی شرح داده شده است، که شامل مراحل زیر است:

جمع آوری اسناد: در ابتدای متن‌کاوی به جمع‌آوری اسناد پرداخته می‌شود.

پیش پردازش: شامل پیش پردازش متون و پیش پردازش ادبی است.

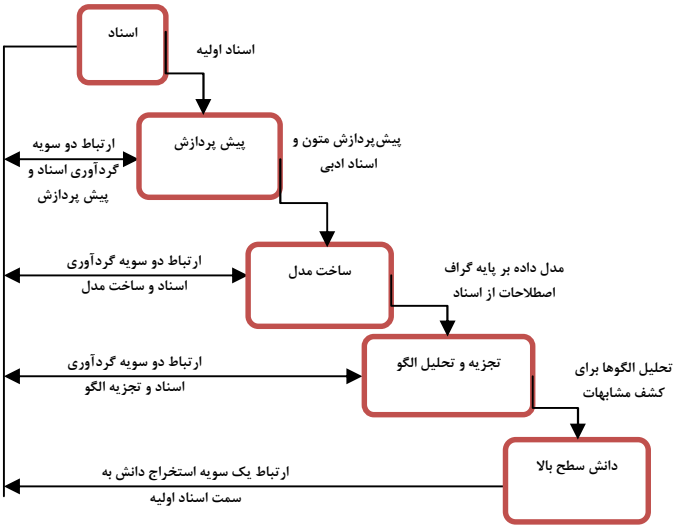
- ✓ **پیش پردازش متون**: بر روی مجموعه‌ی اسناد جمع آوری شده، نشانه‌گذاری[1]، حذف کلمه توقف یافته[2] و ریشه‌یابی[3] انجام می-شود[٥].
- ✓ **پیش پردازش ادبی**: پیش پردازش ادبی، برای افزایش اطلاعات عبارت‌ها بکار رود. بدین منظور دیدگاه‌های، برچسب گذاری ادات سخن[4]، تیکه کردن متن[5]، ابهام زدایی حس کلام[6] و تجزیه کردن[7] را مکرراً اعمال می‌کند[٦].

ساخت مدل: در این مرحله، یک مدل داده‌ای بر پایه گراف اصطلاحات اسناد ساخته و خصوصیات آن‌ها استخراج می‌شوند.

تجزیه و تحلیل الگو: تحلیل الگوها، برای کشف مشابهت‌های بین اسناد بر پایه اصطلاحات مهم اعمال می‌شود که به‌وسیله خوشه‌بندی برای شکل دهی گروه‌های اسناد و استخراج اصطلاحات کلیدی بکار می‌رود.

دانش سطح بالا: فرایندی که به‌وسیله استخراج اصطلاحات کلیدی از خوشه‌ها و تعیین هویت موضوع هر خوشه نتیجه می‌شود[٥].

فرایند متن‌کاوی پیشنهادی این مقاله در شکل ١، نمایش داده شده است. همان‌طور که در این شکل مشاهده می‌شود، از مرحله پیش‌پردازش، ساخت مدل و تجزیه و تحلیل الگو با جمع‌آوری اسناد ارتباط دو سویه بر قرار است. در مرحله آخر، بدست آمدن دانش نهایی، ارتباط یک سویه با اسناد اولیه وجود دارد.



شکل ١: مراحل فرایند متن‌کاوی شامل جمع‌آوری اسناد، پیش پردازش، ساخت مدل، تحلیل الگو و استخراج دانش است که با خط‌های ارتباطی یک سویه و دو سویه با هم در ارتباطند.

٣- متن‌کاوی در یادگیری الکترونیکی

امروزه آموزش الکترونیکی که از روش یادگیری الکترونیکی استفاده می‌کند در مقایسه با روش‌های آموزش واقعی از محبوبیّت قابل توجهی برخوردار شده است. محیط یادگیری الکترونیکی، یک سیستم یادگیری بر خط است که به منظور کمک به مدیریت دروس آموزشی برای یادگیرندگان، طراحی شده است که آن را سیستم مدیریت دروس[8] یا سیستم مدیریت یادگیری[9] نامیده‌اند.

متن‌کاوی در حوزه‌های مختلف آموزش الکترونیکی کاربرد دارد و از آن جمله؛ تهیه خلاصه‌ای از کیفیت دانش یادگیرندگان، تهیه خلاصه فرایندهای یادگیری یادگیرندگان، تهیه خلاصه‌ای از فرایندهای مباحثات

[1] Tokenization
[2] Stop-Word Removal
[3] Stemming
[4] Part-of-Speech tagging (POS)
[5] Text Chunking
[6] Word Sense Disambiguation (WSD)
[7] Parsing
[8] Course Management system (CMS)
[9] Learning Management System (LMS)

Downloaded from jor.iranaict.ir at 0:57 +0330 on Monday November 9th 2026

یادگیرندگان، یافتن یادگیرندگانی که به کمک اساتید نیاز دارند، تجزیه و تحلیل محتوای آموزش الکترونیکی و تجزیه و تحلیل مشخصات هر یادگیرنده در مباحثات می‌توان نام برد. بر اساس موارد ذکر شده در بالا، منافع زیر در یادگیری الکترونیکی فراهم می‌شوند:

- ✓ اساتید می‌توانند فرایندهای یادگیری یادگیرندگان را بشناسند و دانش آن‌ها را به منظور دادن بعضی مقدمات آموزش‌های مؤثر برای یادگیرندگان تعیین کنند.
- ✓ اساتید می‌توانند محتوای یادگیری الکترونیکی را تجزیه و تحلیل کنند و بهبود بخشند.
- ✓ اساتید می‌توانند مباحثاتی را که به طور کلی یادگیرندگان به آن می‌پردازند را به دست آورند.
- ✓ یادگیرندگان می‌توانند کیفیت دانش و فرایندهای یادگیری خود را به وسیله خویش تشخیص دهند[۷].

٤- روش‌های متن‌کاوی

در سال‌های اخیر، محققان شروع به بررسی روش‌های متن‌کاوی برای کمک به ساختارها، مدیران، اساتید و یادگیرندگان کرده‌اند تا سیستم‌های یادگیری الکترونیکی را بهبود بخشند. روش‌های متن‌کاوی به قرار زیر هستند.

استخراج اطلاعات[1]: نقطه شروع برای رایانه‌ها تحلیل متون ساخت نایافته نایافته به منظور استخراج اطلاعات است. نرم افزار استخراج اطلاعات، عبارات کلیدی را استخراج و ارتباطات متون را تعیین هویت می‌کند. این روش می‌تواند بسیار مفید باشد وقتی که با حجم زیاد متون سروکار دارد.

پیگردی عناوین[2] : سیستم پیگردی عناوین برای نگهداری پروفایل و اسناد کاربران، بکار می‌رود و بر اساس آن‌ها، اسناد دیگری برای کاربران پیش بینی می‌کند. این سیستم بر اساس تاریخ خواندن اسناد توسط کاربران استنتاج می‌کند و از میان اطلاعات، اسناد مناسب هر کاربر را انتخاب می‌کند.

خلاصه‌سازی[3]: یکی از روش‌های مفید و قابل توجه، ابزار خلاصه‌سازی متون است. هدف ابزارهای خلاصه‌سازی متون، کاهش طول جزئیات یک سند است در حالی که نکته های اصلی و مفهوم کلی آن‌ها را حفظ می‌کند [۳].

گروه‌بندی[4]: روش گروه‌بندی در متن‌کاوی، عبارت است از تعیین هویت عناوین اصلی که از شمارش کلمه‌ها در اسناد از پیش تعریف شده به دست می‌آیند. ابزار گروه‌بندی، اسناد را بر اساس، بیشترین محتوا روی یک عنوان مخصوص، درجه‌بندی می‌کند.

خوشه‌بندی[5]: روش خوشه‌بندی در متن‌کاوی، گروه کردن اسناد مشابه بر پایه محتوای آن‌ها است [٤]. به طوری که اسنادی که دارای خصوصیات و ویژگی‌های مشابه هستند در یک دسته قرار می‌گیرند.

پیوند مفاهیم[6]: ابزارهای پیوند مفاهیم، اسناد مرتبط را به‌وسیله‌ی تعیین هویت مفاهیم با هم پیوند می‌دهد. ابزارهای پیوند مفاهیم است که ممکن است با استفاده از روش‌های جستجوی سنتی آن‌ها را نیابند. پیوند مفاهیم یک محتوای ارزشمند در متن‌کاوی است[۳].

مصوّرسازی اطلاعات[7]: مصوّرسازی گرافیکی اطلاعات، اطلاعات قابل درک بهتری به‌وسیله متون خالص مبتنی بر توصیف‌ها، فراهم می‌کند. همچنین، این روش، کمک می‌کند تا مجموعه‌های اسناد بزرگ استخراج شوند[٦].

پرسش و پاسخ[8]: حوزه کاربرد دیگر پردازش زبان طبیعی، سؤال‌های زبان زبان طبیعی است که با روش پاسخگوی سؤال‌ها، بهترین پاسخ را می‌یابد.

قوانین انجمنی[9]: قوانین انجمنی روشی است برای کشف ارتباط‌هایی میان میان یک مجموعه بزرگ متغیرها که در مجموعه‌ی داده‌ها استفاده می‌شوند.

٥- مروری بر متن‌کاوی در آموزش الکترونیکی

در این قسمت مروری بر کارهای مرتبط آموزشی با استفاده از روش-های متن‌کاوی در یادگیری الکترونیکی می‌شود.

٥-١- استخراج اطلاعات آموزشی

امروزه با توجه به رشد سریع حجم متون در پایگاه داده‌ها به خصوص در شبکه‌ی جهانی اینترنت، کار استخراج اطلاعات از متون، اهمیت یافته و مورد بررسی محققان قرار گرفته است. اساتید ابزارهای بسیار کمی برای نمایش و پیگیری فعالیت‌های دانشجویان و توانایی کمی برای کشف و حل مشکلات یادگیرندگان در این سیستم آموزشی را دارند. بنابراین، گرفتن

[1] Information Extraction
[2] Topic Tracking
[3] Summarization
[4] Categorization
[5] Clustering
[6] Concept Linkage
[7] Information Visualization
[8] Question Answering
[9] Association Rule

Downloaded from jor.iranaict.ir at 0:57 +0330 on Monday November 9th 2026

تصوری واضح از هر یادگیرنده یا گروه و پیشرفت آن‌ها در دروس، در طول‌ترم سخت و برای استاد زمان‌بر است. اما در همان زمان، این سیستم‌های استخراج اطلاعات مقدار زیاد اطلاعات را ذخیره می‌کنند که بسیار ارزشمند و قابل استخراج است. همچنین می‌توانند برای تحلیل رفتار، پیش‌بینی کارایی، نمره نهایی و گروه‌بندی یادگیرندگان بر حسب اولویت‌های آن‌ها استفاده شوند[۸].

دیدگاه‌هایی که در حوزه استخراج اطلاعات استفاده می‌شوند، شامل تجزیه، تحلیل و دوباره سازی اسناد است. دیدگاه‌های دیگر شامل تعیین هویت و استخراج اطلاعات نیمه ساخت یافته مهم، استخراج کلمات و عبارات کلیدی از اسناد است که نشانه‌گذاری و تطبیق عبارات را استفاده می‌کنند. این روش سهم قابل توجه‌ای را در آموزش الکترونیکی به منظور توانایی در استخراج خودکار اطلاعات مفید، دارا است [۵].

روش استخراج اطلاعات آماری پیوندی با استفاده از جملات bi-gram pseudo، نماینده روش‌های آماری در روش خلاصه‌سازی متون است و در روش استخراج اطلاعات آماری پیوندی پیشنهادی نیز، استفاده می‌شود [۹]. درک زبان طبیعی به طور کامل، هنوز دور از توانایی فناوری فعلی است. وجود روش‌هایی چون استخراج اطلاعات هستند که با دقت مناسب، توانایی شناسایی چندین نوع موجودیت در متون و تعیین برخی ارتباطات بین آن‌ها، زبان طبیعی را بیان می‌کنند [۱۰].

این مقاله[۱۱]، الگوریتمی مورد بحث قرار می‌دهد که چگونگی یافتن صفحات وب مورد درخواست کاربر را با استفاده از روش‌های متن‌کاوی یا چگونگی استخراج و بیان ویژگی‌های متون را نشان می‌دهد. در این مقاله[۱۲]، استخراج ایده‌هایی که برای حل مسائل فنی بکار می‌روند به عنوان فرایند استخراج ایده‌های جدید و مفید از متون ساخت نایافته است.

۵-۲- پیگردی عناوین آموزشی

ردیابی یا پیگردی عناوین، یکی از روش‌های متن‌کاوی است که برای حجم زیاد متون کاربرد دارد. در این مقاله [۱۳]، مسئله کشف عناوین و پیگردی[1]، از یک جریان اسناد متنی پردازش می‌شود. در این مقاله [۱٤] جزئیات ویژگی‌ها، کارایی قطعه‌بندی[2] و تقسیم سیستم‌های پیگردی به وسیله‌ی سیستم‌های اژدها[3] برای کشف عناوین و ارزیابی پیگردی ارائه می‌شوند.

در این مقاله[۱۵]، سیستم پیگردی عناوین مبتنی بر مدل یونی گرام توصیف می‌شود و به‌وسیله سیستم‌های اژدها برای کشف عناوین و ارزیابی پیگردی ارائه شد. این مقاله [۱٦]، خلاصه‌ی وظایف کشف و پیگردی عناوین و نتایج ارزیابی آن‌ها را ارائه می‌کند. همچنین این مقاله [۱۷]، سیستم پیگردی را برای کشف عناوین و پروژه منابع چندگانه اطلاعات در شکل متون، ارائه می‌کند.

۵-۳- خلاصه‌سازی محتوای آموزشی

در چندین سال اخیر با توجه به رشد روز افزون داده و اطلاعات و گسترش منابع اطلاعاتی از جمله اینترنت نیاز مبرمی برای خلاصه‌سازی متون پیدا شده است. در سیستم‌های آموزش الکترونیکی، با کمک نرم افزارهای خلاصه‌سازی می‌توان یادداشت‌ها و نکاتی که یادگیرندگان در انجمن‌های بحث و گفتگو مطرح می‌کنند را خلاصه کرد.

در مقاله [۹]، مقایسه بین روش پیشنهادی با استفاده از اطلاعات متنی و روش‌های دیگر خلاصه‌سازی متون در "با عنوان[4]" و "بدون عنوان[5]" برای خلاصه‌سازی سند مجزا بیان شده است. هدف این مقاله [۱۸]، استخراج و خلاصه کردن ویژگی‌های مشتریان محصولات است.

این مقاله [۱۹]، روش استخراج خلاصه متن‌کاوی را توصیف می‌کند که با فرایند تحلیلی استفاده شده به‌وسیله‌ی تجزیه و تحلیل ثبت اختراع مطابقت می‌کند. در این مقاله[۲۰]، روش تجزیه و تحلیل تقارن معنایی[6] فناوری رایانه برای تحلیل اطلاعات به طور خودکار بکار می‌رود و آنچه که کاربران به آن علاقه دارند را خلاصه می‌کند.

۵-٤- روش‌های دیگر

در این قسمت نیز به دیگر روش‌های متن‌کاوی اشاره شده است که برای حجم زیاد متون و کاربردهای آموزشی بکار برده می‌شوند. امروزه با حجم وسیعی از اطلاعات روبه رو هستیم که قصد ذخیره سازی و نمایش آن‌ها را داریم. از روش‌های مدیریت داده‌ها در سیستم آموزش الکترونیکی می‌توان به خوشه‌بندی یادگیرندگان بر اساس مشخصات مشابه آن‌ها، اشاره کرد [۲۱]. هدف خوشه‌بندی، جداسازی یادگیرندگان بر اساس به‌کارگیری الگوهای استخراج شده از داده‌ی سیستم یادگیری الکترونیکی است که می‌تواند بر حسب شباهت محتوای صفحات وب درسی باشد [۲۲].

در این مقاله [۲۳]، یک روش خوشه‌بندی رمان کاربر[7] به منظور سازماندهی جوامع برای یک مجموعه یادگیرندگان همفکر در یادگیری مشارکتی بر روی سیستم BlogGrid، پیشنهاد می‌شود. در این مقاله [۲٤]، یک روش شناسی برای اندازه‌گیری کیفیت سلسله مراتب خوشه در

[1] Topic Detection and Tracking(TDT)

[2] Segmentation

[3] Dragon System

[4] With Title

[5] Without Title

[6] Semantic Polarity Analysis

[7] Novel User Clustering

Downloaded from jor.iranaict.ir at 0:57 +0330 on Monday November 9th 2026

عبارات اندازه F[1] معرفی می‌شود و نتایج مقایسه‌ی الگوریتم‌های مختلف آزمایش‌ها ارائه می‌شوند.

روش دیگر متن‌کاوی، گروه‌بندی متون است که به کاربر یک دید کلی از آنچه که در متون وجود دارد، می‌دهد. هدف آن، پیش‌بینی برآورد ارزش ناشناخته یک متغیر است که یادگیرنده را توصیف می‌کند. هدف این روش، ایجاد گروه‌های یادگیرندگان بر حسب ویژگی‌های شخصی آن‌ها است [۲۵].

همچنین در مقاله[۲۶]، روش‌های استفاده شده برای انجام متن‌کاوی روی مجموعه‌های اطلاعات متنی مورد بحث و ارزیابی قرار می‌گیرد.

یکی دیگر از روش‌های متن‌کاوی، مصوّرسازی اطلاعات است. هدف تحلیل و مصوّرسازی اطلاعات، برجسته کردن اطلاعات مفید و حمایت از تصمیم‌گیری است. در محیط آموزش الکترونیکی، این روش، می‌تواند به مربیان و اساتید کمک کند تا فعالیت‌های درسی یادگیرنده‌ها را تحلیل کنند و در نتیجه یک دید کلی در مورد یادگیری یادگیرندگان بدست آورند.

همچنین با استفاده از روش پیوند مفاهیم، ابزارهایی در یادگیری الکترونیکی وجود دارند که اسناد مرتبط را به وسیله‌ی تعیین هویت مفاهیم با یکدیگر پیوند می‌زنند.

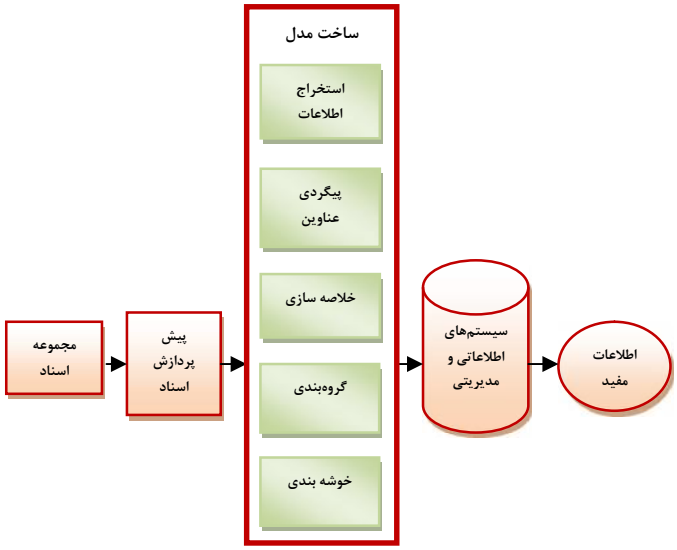
روش کاربردی دیگر متن‌کاوی، پاسخ‌گوی سؤال است. چنین روشی در آموزش الکترونیکی نیز کاربرد دارد به این صورت که سؤال‌های یادگیرندگان در موارد گوناگون مطرح شود و سپس توسط اساتید/مدیران پاسخ داده شوند.

در روش قوانین انجمنی، هدف، فراهم کردن بازخورد برای حمایت مدیر، اساتید و نویسندگان دروس در تصمیم گیری‌ها است (در مورد این که چطور یادگیری یادگیرندگان و سازماندهی منابع آموزشی مؤثرتر، بهبود یابند).

۶- ارائه راهکاری بر اساس مدل پیشنهادی

هدف اولیه‌ی روش استخراج اطلاعات، تعیین هویت داده‌ها و ارتباطات بین آن‌ها در متون است. الگوریتم‌های استخراج اطلاعات ممکن است لغتنامه‌هایی را برای تعیین هویت برخی شرایط و الگوهای ادبی بکار برند. این روش از پتانسیل بالایی در آموزش الکترونیکی به واسطه توانایی استخراج خودکار اطلاعات مفید برخوردار است. همچنین طبق بررسی‌ها و مطالعات انجام شده روش‌های همچون پیگردی عناوین، خوشه‌بندی، گروه‌بندی و خلاصه‌سازی متون نسبت به دیگر روش‌های متن‌کاوی کاربرد بیشتری در یادگیری الکترونیکی جهت استخراج اطلاعات دارند.

جهت کاربردی‌تر کردن فرایند متن‌کاوی در حوزه یادگیری الکترونیکی، راهکاری بر اساس مدلی که در شکل ۳ آمده است، پیشنهاد می‌شود. این مدل پیشنهادی فرایند متن‌کاوی را به منظور استخراج اطلاعات (دانش) نشان می‌دهد.



شکل ۳: ترکیب روش‌های متن‌کاوی کاربردی‌تر در فرایند متن‌کاوی جهت استخراج اطلاعات

با توجه به شکل ۳، در می‌یابیم که، اولین مرحله در مدل پیشنهادی عبارت است از گردآوری اسناد. در این مرحله متون و اسناد ادبی برای انجام فرایند متن‌کاوی جمع‌آوری می‌شود. سپس بررسی مجموعه اسناد گردآوری شده در واحد "پیش‌پردازش اسناد" به منظور پیش پردازش قالب و ویژگی‌های آن اسناد انجام می‌گیرد. این مرحله پیش‌پردازش به صورت زیر صورت می‌گیرد:

- نشانه‌گذاری: تعیین هویت تمام کلمه‌ها (فرایند شکستن یک جریان متن به کلمه‌ها، عبارت‌ها، نمادها و دیگر عناصر معنی‌دار به نام نشانه) است.
- حذف کلمه توقف یافته: کلمه‌های مکرر مانند، "و"، "یک" و "یا" حذف می‌شوند.
- ریشه‌یابی: کاهش شکل‌های مختلف یک کلمه به یک شکل مجزاء است.
- برچسب‌گذاری ادات سخن: ماهیت کلمات عبارات را مشخص می‌کند. مانند، اسم، فعل و صفت برای هر عبارت.
- ابهام‌زدایی حس کلام: برای حل و فصل ابهام در معنای تک واژه‌ها یا عبارات تلاش می‌کند.

[1] FMeasure

Downloaded from jor.iranaict.ir at 0:57 +0330 on Monday November 9th 2026

- تجزیه کردن: یک درخت تجزیه کامل جمله را تولید می‌کند. در نتیجه تجزیه، می‌توان ارتباط هر کلمه در جمله با دیگر کلمات را پیدا کند.

در قدم بعد خروجی واحد"پیش‌پردازش اسناد" به واحد "ساخت مدل" داده می‌شود. ۵ روش، استخراج اطلاعات، پیگردی عناوین، خوشه-بندی، گروه‌بندی و خلاصه‌سازی متون برای ساخت و اجرای مدل اعمال می شوند. این ترکیب روش‌های متن‌کاوی واحد "ساخت مدل" نسبت به دیگر ترکیبات روش‌های متن‌کاوی کاربرد بیشتری دارد. بدست آمدن این ترکیب بر اساس مطالعات انجام شده در این روش‌ها بدست آمده است. هنگامیکه قصد بر آن است که مدل ساخته شود، می‌توان از هر کدام این روش‌ها به تنهایی یا ترکیبی از آن‌ها استفاده شود. همچنین تمام نکاتی که در تعاریف این روش‌ها بیان شد، رعایت می‌شود. بنابراین مراحل ساخت مدل به صورت زیر انجام می‌گیرند:

- ارائه سند (فضای گراف اسناد)
- استخراج خصوصیات اصطلاحات اسناد (بر اساس روش‌های متن-کاوی ذکر شده در شکل ۳)

نتایج واحد ساخت مدل، می‌توانند در واحد "سیستم‌های اطلاعاتی و مدیریتی" قرار گیرند و تجزیه و تحلیل شوند. مراحل این قسمت بدین صورت هستند:

- محاسبه شباهت‌ها
- استخراج اصطلاحات کلیدی

در خروجی اطلاعات مفیدی شامل، خوشه‌های اسناد، عناوین خوشه‌ها و اصطلاحات کلیدی استخراج شده، در قالب دانش نهایی نتیجه می شوند.

با بکارگیری شکل ۳، در سیستم یادگیری الکترونیکی به صورت زیر عمل می‌شود:

- ابتدا به جمع‌آوری داده پرداخته می‌شود. سیستم یادگیری الکترونیکی به وسیله یادگیرندگان استفاده می‌شود و اطلاعات آن‌ها در پایگاه داده سیستم ذخیره می‌شود.
- پیش پردازش داده: داده پاک می‌شود و به یک شکل مناسب استخراج شده، انتقال می‌یابد. به منظور پیش پردازش داده سیستم مدیریت یادگیری، ابزار مدیر پایگاه داده یا بعضی ابزارهای پیش پردازش مخصوص را به کار می‌برد.
- اعمال روش‌های متن‌کاوی به منظور ساخت مدل: روش‌های متن-کاوی برای ساختن و اجرا کردن مدلی اعمال می‌شود که دانش مورد علاقه کاربران (اساتید، یادگیرندگان و مدیران) را کشف و خلاصه می‌کند.
- تجزیه و تحلیل و استقرار نتایج: نتیجه و مدل به دست آمده توسط اساتید تفسیر شده و برای عملیات بیشتری به کار می‌رود. اساتید می‌توانند اطلاعات مکشوف را برای تصمیم‌گیری در مورد فعالیت‌های درسی یادگیرندگان به منظور بهبود یادگیری آنان به کار برند.

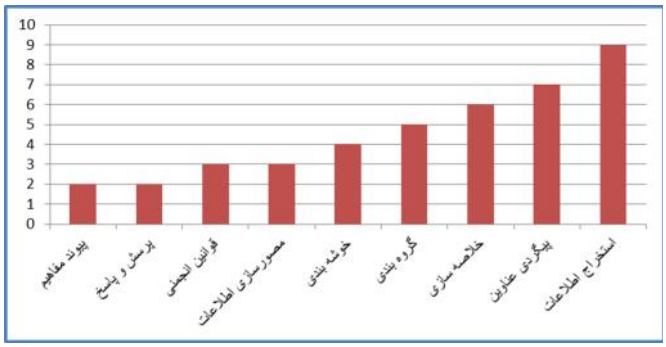
با استفاده از روش‌های متن‌کاوی، مسیرهایی برای اساتید فراهم می شود تا دروس بر خط را تولید کنند و با استفاده از آن‌ها یادگیرندگان را به یادگیری الکترونیکی مؤثرتری رهنمون کنند.

شناخت در مورد یادگیرندگان، یکی از مفاهیم کلیدی برای موفقیت آموزش بر خط است. چه انگیزه‌هایی، یادگیرندگان در محیط یادگیری بر خط دارند؟ چه چیز آن‌ها از اساتید یاد می‌گیرند و چه نمی‌آموزند؟ این‌ها سؤال‌های مهمی به منظور شکل‌دهی فعالیت‌های آموزشی هستند. در محیط‌های یادگیری سنتی، اساتید توانایی دیدن یادگیرندگان را به طور حضوری دارند. اما، این فرصت در محیط یادگیری الکترونیکی وجود ندارد. یکی از راه‌های قابل اجرا، جمع‌آوری اطلاعات اداره نظرسنجی و پرسشنامه از یادگیرندگان است. این کاربرد سؤال برانگیز است و ممکن نیست به هیچ وجه طبیعت درست یادگیرندگان را منعکس کند. تعامل یادگیرندگان در محیط‌های آموزش الکترونیکی سرنخ‌های مختلفی در مورد ویژگی‌های یادگیرندگان می‌دهد.

با به‌کارگیری روش‌های متن‌کاوی برای یادگیری همکارانه در سیستم آموزش الکترونیکی، نتایج زیر حاصل شدند:

- کشف یادگیرندگانی که فرایندهای یادگیری بی قاعده و بدون اصولی در آموزش الکترونیکی دارند.
- توانایی تجزیه و تحلیل مفاهیم و مفاد درسی بر خط برای یادگیری الکترونیکی
- توانایی تجزیه و تحلیل و ثبت داده یادگیری یادگیرندگان با استفاده از درخت تصمیم
- به دست آوردن آنتروپی (واحد اندازه‌گیری بی نظمی در یک مجموعه) بر پایه تجزیه و تحلیل‌ها در تابلو مباحثات یادگیرندگان

در شکل ٤، نمودار فراوانی مقالاتی را نشان می‌دهد که در مورد روش های متن‌کاوی بحث و مطالعه شده است.



شکل ٤: نمودار فراوانی مقالات در مورد روش‌های متن‌کاوی

Downloaded from jor.iranaict.ir at 0:57 +0330 on Monday November 9th 2026

۷- نتیجه گیری و مطالعات آتی

هدف از این مقاله، فراهم آوردن نگرشی بر روش‌های متن‌کاوی در آموزش الکترونیکی و بررسی آن‌ها است. متن‌کاوی به عنوان کشف دانش در متون شناخته می‌شود و عموماً به فرایند استخراج اطلاعات و دانش مورد نظر از متون ساخت نایافته اشاره می‌کند. مسئله کشف دانش از متون، استخراج مفاهیم صریح و ضمنی و ارتباط معنایی بین مفاهیمی است که از روش‌های پردازش زبان طبیعی استفاده می‌کنند.

روش‌های متن‌کاوی به منظور بهبود یادگیری و آموزش در ارتباطات مجازی، ممکن است مورد استفاده قرار گیرند. متن‌کاوی در این مسیر، روش‌هایی را برای آموزش مربوط به یادگیرندگان به منظور تصمیم گیری‌های مؤثر در مورد روش تدریس، مدل پیشرفت و سرمایه‌گذاری در زیرساخت، فراهم می‌آورد.

اخیراً، متن‌کاوی آموزشی به یک زمینه تحقیقاتی در حال ظهور تبدیل شد که برای استخراج دانش و کشف الگوها از سیستم آموزش الکترونیکی مورد استفاده هستند. این سیستم آموزشی در حال حاضر پیش روی تعدادی از مسائل از قبیل شناسایی نیازهای یادگیرندگان، آموزش شخصی و پیش‌بینی کیفیت تعاملات یادگیرنده‌ها است. متن‌کاوی آموزشی مجموعه‌ای از روش‌ها را فراهم می‌کنند، که می‌تواند کمک کند تا نظام آموزشی بر این مسائل غلبه کند. کار متن‌کاوی آموزشی اجازه شناسایی و موقعیت‌یابی اطلاعات برای پردازش‌های آموزش الکترونیکی که نیاز به بهبود دارند، یا آن‌ها که خیلی خوب کار انجام می‌دهند و می‌توانند به عنوان نمونه‌های خوب استفاده شوند، است. متن‌کاوی آموزشی در این زمینه امکان تجزیه و تحلیل، بهتر فهمیدن و یاد گرفتن فرایندهای آموزشی را با استفاده از روش‌های متن‌کاوی، بررسی می‌کند. این کار به طور کلی امکان تمرکز روی عملکرد سیستم مدیریت یادگیری و استخراج فرایند مدل داده‌ای را می‌دهد. به علاوه، اجازه شناسایی راه‌های مؤثر بیشتری برای فرایند تدریس را می‌دهد که می‌تواند به منظور ارتقاء فرایندهای آموزشی مورد استفاده قرار گیرند.

همچنین متن‌کاوی در حوزه‌های زیر در آموزش الکترونیکی کاربرد دارد:

- تهیه خلاصه‌ای از کیفیت دانش یادگیرنده‌ها
- تهیه خلاصه فرآیندهای یادگیری یادگیرنده‌ها
- تهیه خلاصه‌ای از فرآیندهای مباحثات یادگیرنده‌ها
- پیش بینی کیفیت دانش یادگیرنده‌ها در آینده
- یافتن یادگیرنده‌هایی که به کمک اساتید نیاز دارند.
- تجزیه و تحلیل محتوای آموزش الکترونیکی

برای محورهای مطالعه آتی می‌توان، گسترش روش‌های یادگیری نیمه نظارتی برای استخراج اطلاعات، ساختن لغت‌نامه‌های رضایتمند با روش‌های بهتر، متن‌کاوی یک چهارچوب مجزا که هر دو داده ساخت یافته و ساخت نیافته‌ی یک دامنه مخصوص برای استخراج دانش کامل‌تر هستند و پردازش خودکار پیغام‌های تکنیکی روی تابلو بحث، ایمیل و لاگ‌های پیغام مستقیم را، نام برد.

مراجع

[1] Cristobal Romero, Sebastian Ventura, Enrique Garcia; "Data mining in course management systems: Moodle case study and tutorial", Spain, Computers & Education 51 (2008) 368–384.

[2] Cristóbal Romero, Sebastián Ventura, Senior; "Educational Data Mining: A Review of the State-of-the-Art" Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEETransactions on Issue Date,,VOL: 40 Issue:6,On page(s): 601 – 618, 2010.

[3] Vishal Gupta, Gurpreet S. Lehal; " A Survey of Text Mining Techniques and Applications" Journal of emerging technologies in web intelligence,VOL. 1, NO. 1, Aug.2009.

[4] Jong Hwan Suh, Chung Hoon Park, Si Hyun Jeon; "Applying text and data mining techniques to forecasting the trend of petitions filed to e-People" Information and Communication Technology Research and Development Center, Expert Systems with Applications 37, 7255–7268, 2010.

[5] Khaled Hammouda1 and Mohamed Kamel; "Data Mining in e-Learning", E-Learning Networked Environments and Architectures Advanced Information and Knowledge Processing, Springer, Part IV, 374-404,2007.

[6] Andreas Hotho, Andreas N¨urnberger, Gerhard Paaß; "A Brief Survey of Text Mining" machine learning, May 2005.

[7] Maomi Ueno; "Data mining and text mining technologies for collaborative learning in an ILMS "Samurai" Nagaoka University of Technology,on Advanced Learning Technologies, In IEEE international Conference on Advanced Learning Technologies, Washington, DC,1052-1053, 2004.

[8] Marta Zorrilla, Diego García,Elena Álvarez; "A Decision Support System to improve e-Learning Environments" Lausanne, Switzerland ,ACM, EDBT 2010, March 22-26, 2010.

[9] Youngjoong Ko, Jungyun Seo," An effective sentence-extraction technique using contextual information and statistical approaches for text summarization", Available online 4 March 2008, Pattern Recognition Letters 29 (2008) 1366–1371.

[10] Raymond J. Mooney,Razvan Bunescu,"Mining Knowledge from Text Using Information Extraction", Department of Computer Sciences,University of Texas at Austin,1 University Station C0500, Austin, TX 787120233,2005.

[11] Shiqun Yin, Yuhui Qiu, Jike Ge; " Research and Realization of Text Mining Algorithm on web", International Conference on Computational Intelligence and Security Workshops, CISW,on ,413 - 416 ,15-19 Dec. 2007.

[12] Dirk Thorleuchter,Dirk Van den Poel,Anita Prinzie,"Mining ideas from textual information", Expert Systems with Applications 37 (2010) 7182–7188.

[13] K. Rajaraman, Ah-Hwee Tan,"Topic Detection, Tracking and Trend Analysis Using Self-organizing Neural Networks", Kent Ridge Digital Labs,21 Heng Mui Keng Terrace, Singapore 119613,2001.

[14] P. van Mulbregt, I. Carp, L. Gillick, S. Lowe and J. Yamron, "Text Segmentation and Topic Tracking on Broadcast News via a Hidden Markov Model Approach", Dragon Systems, Inc,320 Nevada, Street Newton, MA 02460,1998.

[15] J.P. Yamron, I. Carp, L. Gillick, S. Lowe, and P. van Mulbregt,” Topic Tracking in a News Stream”, Dragon Systems, Inc. 320 Nevada Street, Newton, MA 02460,1999.

[16] Jon Fiscus, George Doddington, John Garofolo, Alvin Martin, "NIST’S 1998 Topic Detection and Tracking Evaluation (TDT2) ", National Institute of Standards and Technology, MD 20899-8940 USA,Aug 2007.

[17] Hubert Jin, Rich Schwartz, Sreenivasa Sista, Frederick Walls, "Topic Tracking for Radio, TV Broadcast, and Newswire", BBN Technologies70 Fawcett Street, Cambridge,MA, 02138. Cambridge, MA 02138,Apr 1999.

[18] Minqing Hu,Bing Liu,” Mining and Summarizing Customer Reviews”, Department of Computer Science,University of Illinois at Chicago,851 South Morgan Street, Chicago, IL 60607-7053,2009.

[19] Yuen-Hsien Tseng, Chi-Jen Lin, Yu-I Lin,”Text mining techniques for patent analysis”, Available online 26 January 2007, Information Processing and Management 43 (2007) 1216–1247.

[20] Li Yu, Qiang Li,”A Novel Web Text Mining Method based on Semantic Polarity Analysis”,2009.

[21] Amelia Zafra, Cristóbal Romero, Sebastián Ventura,” Multiple instance learning for classifying students in learning management systems”, Department of Computer Science and Numerical Analysis, University of Cordoba, Spain, Expert Systems with Applications 38 (2011) 15020–15031.

[22] Ioannis Kazanidis, Stavros Valsamidis, Theodosios Theodosiou, Sotirios Kontogiannis," PROPOSED FRAMEWORK FOR DATA MINING IN E-LEARNING: THE CASE OF OPEN E-CLASS" IADIS International Conference Applied Computing 2009, Democritus University of Thrace, 2009.

[23] Jason J. Jung,” Social grid platform for collaborative online learning on blogosphere:A case study of eLearning BlogGrid”, Department of Computer Engineering, Yeungnam University, Dae-Dong, Gyeongsan, Republic of Korea, Expert Systems with Applications 36 (2009) 2177 2186.

[24] Bjornar Larsen and Chinatsu Aone,"Fast and Effective Text Mining Using Linear-time Document Clustering", SRA International, Inc. 4300 Fair Lakes Cow-l, Fairfax, VA 22033,1999.

[25] Fu-Ren Lin, Lu-Shih Hsieh, Fu-Tai Chuang,” Discovering genres of online discussion threads via text mining”, Computers & Education 52 (2009) 481–495.

[26] ursun Delen *, Martin D. Crossland,” Seeding the survey and analysis of research literature with text mining”, Department of Management Science and Information Systems, Expert Systems with Applications 34 (2008) 1707–1720.

Downloaded from jor.iranaict.ir at 0:57 +0330 on Monday November 9th 2026