

نقش تحلیل متن و بازیابی اطلاعات در حوزه حسابداری

سید علی حسینی^۱، صغری فصیحی^۲، نسیم محمودیان^۳

تاریخ دریافت: ۹۴/۰۷/۰۱

تاریخ پذیرش: ۹۴/۱۰/۱۴

چکیده

امروزه بخش زیادی از اطلاعات مالی از طریق اسناد متنی منتشر می‌شود و گروه‌های مختلف از این اطلاعات مکتوب برای اتخاذ تصمیم‌های مالی استفاده می‌نمایند. تاکنون تحقیقات زیادی در دنیا در حوزه تحلیل و بازیابی اسناد و متون حسابداری که می‌تواند بر میزان درک استفاده‌کنندگان از این اطلاعات تأثیر داشته باشد صورت گرفته است. با توجه به مشکلات استفاده‌کنندگان از گزارش‌های مالی قابلیت دسترسی و قابلیت درک اطلاعات و مسائل مرتبط با بازیابی اطلاعات به‌عنوان اهداف اصلی در این مطالعه مورد استفاده قرار گرفته است. بنابراین این مقاله ادبیات تحقیق مرتبط با دو حوزه اصلی مربوط به تحلیل متن و بازیابی اطلاعات در حوزه حسابداری را به شرح زیر بررسی می‌کند: (۱) تحلیل محتوی و قابلیت خوانایی اطلاعات مالی ارائه‌شده و فعالیت‌های متن‌کاوی مربوط به آن و (۲) روند ادبیات تحقیق در ارتباط با بازیابی اطلاعات که دربرگیرنده استخراج متن و نیز مقادیر ارائه‌شده در متن اسناد حسابداری است. مشکل عدم وجود واژگان یکنواخت مسائل مربوط به قابلیت دسترسی و قابلیت درک اطلاعات را تحت تأثیر خود قرار می‌دهد. در آخر تلاش می‌شود تا یک نقشه راه پیشنهادی جهت انجام تحقیقات آتی در حوزه تحلیل متون حسابداری و بازیابی اطلاعات ارائه گردد که شامل لزوم انجام تحقیقات آتی در حوزه‌های (۱) میزان موفقیت تدوین‌کنندگان استاندارد در دستیابی به اهداف مدنظر خود (۲) ساخت واژه‌نامه‌های حسابداری، (۳) شناسایی واژه‌های جدید واردشده به واژه‌نامه و (۴) استانداردسازی واژگان حسابداری است.

واژه‌های کلیدی: تحلیل متن، بازیابی متن، خوانایی

طبقه‌بندی موضوعی: M49

۱ استادیار گروه حسابداری دانشگاه الزهراء (س)، (hosseinira@yahoo.com)

۲ دانشجوی دکتری حسابداری دانشگاه الزهراء (س)، نویسنده مسئول، (Sfasihi2020@yahoo.com)

۳ دانشجوی دکتری حسابداری دانشگاه الزهراء (س)، (nasim_mahmoudian@yahoo.com)

مقدمه

در دهه ۱۹۶۵ میلادی، گلدبرگ (۱۹۶۵) ادعا کرد که «این موضوع مبالغه نیست که بیان کنیم مشکل برقراری ارتباط یک مشکل اساسی در حسابداری است». اسناد متنی حسابداری شامل اصول عمومی پذیرفته شده حسابداری، صورت‌های مالی سالانه شرکت‌ها، به همراه سایر گزارش‌های حسابداری، در عمل ابزارهایی برای انتقال اطلاعات حسابداری محسوب می‌شوند. این اسناد بر طبق استانداردهای پذیرفته شده حسابداری تهیه و منتشر می‌شوند و گروه‌های مختلف از این اطلاعات مکتوب برای اتخاذ تصمیم‌های مالی مناسب بهره می‌گیرند. به هر حال، زبان حسابداری استفاده شده در ارتباط با این منابع اطلاعاتی همواره در حال تغییر است و در عمل این موضوع باعث شده افراد غیرمتخصص به سختی با اصلاحات اخیر انجام شده در حوزه واژگان و مفاهیم حسابداری همراه شوند. به همین دلیل انجمن حسابداران رسمی آمریکا فشار زیادی بر تدوین کنندگان استانداردهای حسابداری وارد کرده است و «تدوین کنندگان استانداردهای حسابداری باید هر بار که یک استاندارد جدید تدوین می‌شود، از معرفی اصطلاحات جدید خودداری نمایند» (بورک، ۱۹۹۷). بعلاوه امروزه اینترنت، منبع مهمی از اسناد الکترونیک برای تحلیل کامپیوتری است. هر شرکت بزرگ، وب‌سایتی دارد که شامل اسنادی علاوه بر گزارش‌های سالانه، کتابچه‌های عرضه عمومی اولیه سهام و گزارش‌های مطبوعاتی است. این اسناد (مانند گزارش‌های تحقیق و توسعه، پست‌های فرصت‌های استخدام، اسناد روابط سرمایه‌گذار و...) می‌تواند شامل اطلاعاتی با ارزش برای تصمیم‌گیرندگان اقتصادی باشد. محققان اقتصاد اطلاعات می‌توانند تحلیل‌های متنی و فن‌های بازیابی اطلاعات را برای مطالعات پیش‌بینی درآمدها و بازده‌ها و همچنین برای مطالعات رخدادهای اقتصادی مانند تقلب، ورشکستگی و ادغام و مالکیت، بکار گیرند. دیگر اسناد متنی مورد توجه برای محققان حسابداری که به‌طور الکترونیکی برای تحلیل‌های متنی و مطالعات اقتصادی در دسترس هستند شامل موارد زیر است: صورت‌های حسابداری، مقررات بورس اوراق بهادار، قوانین مالیاتی، پرونده‌های قضایی مالیاتی، نامه‌های ارسالی به بورس، مقالات تحقیقی و مقالات کاربردی حسابداری و مواد درسی حسابداری (رئوس درسی، شرح برنامه و غیره).

هیئت استانداردهای حسابداری مالی (۲۰۰۵) مشکل قابلیت درک اطلاعات حسابداری را به همراه مشکل ناکافی بودن اطلاعات در نظر گرفته و بازیابی غیر مؤثر اطلاعات را به

استانداردهای حسابداری مالی نسبت می‌دهد. در واکنش به این شرایط، هیئت استانداردهای حسابداری مالی «پروژه تدوین و بازیابی» را برای تدوین ادبیات حسابداری معتبر آغاز کرد. این پروژه در تاریخ ۱ جولای ۲۰۰۹ اجرا شد تا به این ترتیب یک منبع معتبر برای اصول عمومی پذیرفته شده حسابداری غیردولتی در آمریکا ایجاد شود.

تعداد کمی از تحقیقات در حوزه اثربخشی نحوه انتقال رسمی اطلاعات حسابداری از طریق دستورالعمل‌هایی مثل استانداردهای حسابداری مالی صورت گرفته‌اند و این موضوع به طور خاص در ارتباط با حوزه‌های زبان‌شناسی و حسابداری مشاهده می‌شود. با توجه به شکاف پژوهشی در این حوزه، این پژوهش به بررسی ادبیات تحقیق در حوزه تحلیل متن و بازیابی اطلاعات در حوزه حسابداری می‌پردازد. ادبیات تحلیل متن و بازیابی اطلاعات به دو روند تحقیق مجزا قابل تفکیک هستند. اولین روند شامل تحلیل مفاهیم قانونی و محاسباتی روش‌های گزارشگری حسابداری، مطالعات مرتبط با خوانایی اطلاعات حسابداری و فعالیت‌های متن‌کاوی مربوط به آن است. این حوزه به طور مشخص بر استفاده از اجزای حسابداری به منظور ارائه مفاهیم استنتاجی مناسب تمرکز می‌کند (برای مثال ارائه اطلاعات مناسب در ارتباط با عملکرد آتی و قیمت سهام). روند دوم از ادبیات بازیابی اطلاعات تشکیل شده است که هر دو اجزای مربوط به متن و نیز مقادیر بیان شده در متن اسناد حسابداری را دربرمی‌گیرد. این روند بر اهمیت درک واژگان حسابداری و نقش یک فرهنگ لغت جامع حسابداری در بازیابی اطلاعات از اسناد حسابداری الکترونیکی تأکید می‌کند.

قابلیت دسترسی و قابلیت درک

بررسی تاریخی از پژوهش‌های در این حوزه نشان‌دهنده رشد مشخص تحلیل متن و بازیابی اطلاعات در مطالعات تجاری نسبت به دهه‌های گذشته است. این مطالعات از یک مقاله در دهه ۵۰ و دو مقاله در دهه ۶۰ به تعداد ۳۲ مقاله در دهه ۹۰ و ۶۸ مقاله در دهه اول ۲۰۰۰ رسیده است. به این ترتیب پیشرفت‌های انجام شده در فناوری‌های کامپیوتر (در هر دو بخش سخت‌افزار و نرم‌افزار) و افزایش کاربرد روش‌های مختلف ارائه اطلاعات حسابداری در عمل توانسته توسعه انجام شده در این بخش را توضیح دهد (جونز و شوماخر، ۱۹۹۴) و این انتظار وجود دارد که در ادامه فناوری‌های زبان‌شناسی محاسباتی در مطالعات حسابداری و جاری مبتنی بر متن باگذشت زمان بیشتر از قبل مورد استفاده و توجه قرار گیرد.

قابلیت دسترسی و قابلیت درک، همچنان تمرکز اصلی مطالعات تحلیل و بازیابی اطلاعات متن است. تحقیق روی بازیابی متن حسابداری و مقادیر حسابداری گزارش‌های مالی متنی، مسئله قابلیت دسترسی را از چشم‌انداز برنامه‌ریزی، بررسی می‌کند. قابلیت درک گزارش‌های مختلف حسابداری در تحقیقات خوانایی بررسی شده است؛ در ادامه به بررسی تحلیل محتوا و خوانایی اسناد کسب و کار و تحقیقات خوانایی مرتبط پرداخته می‌شود.

تحلیل محتوا و خوانایی اسناد کسب و کار

تحلیل محتوا و تحلیل خوانایی، دو حوزه مرتبط اما به‌خوبی تعریف نشده از مطالعات مبتنی بر متن هستند. جونز و شوماخر (۱۹۹۴) تحلیل‌های موضوعی تجربی (مطالعه روندهای خاص، نگرش‌ها، یا طبقه‌بندی محتوا از متن و سپس استنباط از آن‌ها) و مطالعات ترکیبی را (مطالعه خوانایی متن) را به‌عنوان دو جریان مجزا از تحلیل محتوا، دسته‌بندی می‌کند. کلاتورثی و جونز (۲۰۰۱)، تحقیق دانشگاهی درباره رویه‌های حسابداری را به (۱) مطالعات تحلیل محتوا، (۲) تحقیق خوانایی، تقسیم می‌کند. کلاتورثی و جونز (۲۰۰۶)، همچنین بین تحلیل محتوایی که (۱) معنی محور و متمرکز بر مضامین زیربنایی است و (۲) فرم محور و متمرکز بر ویژگی‌های متنی است، تمایز ایجاد نموده‌اند. این بخش، به‌طور مجزا، تحلیل محتوا و تحلیل خوانایی را با تأیید ریشه‌های مشترک آن‌ها، بررسی می‌کند.

مطالعات تحلیل محتوا در حسابداری

مطالعات تحلیل محتوی در حوزه حسابداری به ۳ صورت (۱) تحلیل محتوای دستی (۲) تحلیل محتوای نیمه کامپیوتری و (۳) تحلیل محتوای کامپیوتری صورت گرفته است.

تحلیل محتوای دستی

تحقیقات گذشته مبتنی بر متن، به‌طور قابل توجهی مبتنی بر تحلیل محتوای دستی که محققان برای استخراج استنباط‌های خود، اسناد متنی را بررسی می‌کنند. در اولین پژوهش‌های این حوزه، محققان به‌طور دستی به تحلیل محتوای گزارش‌های سالیانه برای بررسی استراتژی‌های شرکت و منطق علی می‌پرداختند. به‌طور مشابه، مطالعات حسابداری و مالی، تحلیل محتوا را برای استنباط از افشاهای اختیاری و اطلاعات منتشرشده توسط مطبوعات، پیش‌بینی‌ها و نظرات به‌دست‌آمده مدیریت، مقالات مجلات، کدهای اخلاقی حسابداری حرفه‌ای، یادداشت‌های

همراه صورت‌های مالی، بروشورها، صحبت‌های روسا با سهام‌داران، بحث‌ها و تحلیل‌های مدیریت و از دیگر بخش‌های گزارش سالیانه بکار می‌گیرند. نتایجی که از این کار به دست می‌آید مرتبط با شرکت‌های کوچک‌تر و با سوددهی کمتر است که اطلاعات محدودی را منتشر می‌کنند، شرکت‌هایی که به‌احتمال بیشتری اطلاعات مثبت را منتشر می‌کنند؛ مدیریت تمایل به انتساب پیامدهای منفی به عوامل غیرقابل کنترل دارد درحالی‌که پیامدهای مثبت را به داخل منتسب می‌داند و شرکت‌ها، عمدتاً اطلاعات بیشتری را در دوره‌های افزایش دستاوردهایشان منتشر می‌کنند.

مرکل دیویس و برنن (۲۰۰۷) ادبیات افشای اختیاری اسناد تجاری را بررسی و تحلیل نموده و سه طبقه‌بندی برای طبقات ادبیات موضوع را ایجاد نموده‌اند: انگیزش (افشاسازی‌هایی که مدیریت ادراکات را شکل داده و آن‌هایی که اطلاعات را به‌طور تدریجی می‌افزاید)، چشم‌انداز تحقیق (تهیه‌کننده در مقابل کاربر) و گروهی از استراتژی‌های افشای هفت‌گانه. مدیریت ادراکات، انتقال منتخب اطلاعات است، بطوریکه درک خواننده را از دستاوردهای شرکت، تحریف می‌کند. آن‌ها به این نتیجه رسیدند که رسیدن به نتایج قانع‌کننده‌ای که قابل انتساب به این واقعیت باشد که اغلب مطالعات، یا مدیریت ادراکات یا اطلاعات تدریجی را بررسی می‌کنند و فقط ۸ تا از آن‌ها به هر دو پرداخته‌اند، مشکل است. بعلاوه، اکثریت قریب‌به‌اتفاق مطالعات، یک چشم‌انداز آماده‌کننده دارند. آن‌ها، گزارش‌های را از بیش از یک چشم‌انداز، ارائه می‌کنند.

برنان و همکاران (۲۰۰۹)، ادبیات موضوع مدیریت ادراکات را به‌عنوان پایه‌ای برای تحلیل محتوای دستی ۲۱ عدد از مطبوعات سالیانه در انگلستان بررسی نمودند. آن‌ها استدلال نمودند که تکنیک‌های مدیریت ادراکات ذاتاً ظریف و پیچیده هستند و بنابراین، نیازمند توجه مفصل همراه با تحلیل محتوای دستی از سوی فرد هستند. آن‌ها همچنین بر این باورند که تحلیل جامع‌تر مدیریت ادراکات از طریق کاربرد روش‌ها و عوامل چندگانه، لازم است.

تحلیل محتوای نیمه کامپیوتری

تحلیل محتوای دستی می‌تواند بیانگر ذهنیت محقق باشد. بعلاوه، هزینه‌های بالای پردازش دستی، اندازه نمونه قدرت آزمون‌ها را محدود نموده و قابلیت تعمیم نتایج را کاهش می‌دهد.

میلر (۲۰۰۲) اذعان می‌کند که یک نمونه بزرگ‌تر می‌تواند به تقویت قدرت آزمون‌های آماری و کاهش تأثیر اطلاعات مختل‌کننده، کمک کند، اما داده‌های افشا که به صورت دستی از اسناد متنی جمع‌آوری شده، هزینه‌های بالایی روی نمونه‌های بزرگ، تحمیل می‌کند. با استفاده از نرم‌افزار برای پردازش اسناد متنی الکترونیکی و احتساب فراوانی‌های کلمات، تحلیل محتوای کامپیوتری بر محدودیت‌های مطرح‌شده در تحلیل محتوای دستی، غلبه می‌کند. استنباط نظام‌مند از فراوانی‌های لغوی، به وسیله ادبیات زبان‌شناسی کامپیوتری حمایت شده است. تحلیل آماری فراوانی‌های لغوی، برای حل اختلافات تألیف درباره نوشته‌های مهم تاریخی، بکار گرفته شده است. برای مثال، شواهد فراوانی لغت را به ۱۲ مقاله جیمز مادیسون که در ۱۷۸۸ برای ترغیب شهروندان نیویورک به تصویب قانونی اساسی ایالات متحده نوشته شده، نسبت می‌دهند.

مطالعات حسابداری متنی اندکی، از تحلیل محتوای نیمه کامپیوتری استفاده می‌کنند. مطالعات انجام گرفته توسط اینگرام و فریزر (۱۹۸۳)، فریزر و دیگران (۱۹۸۴) و تینسون و دیگران (۱۹۹۰) بر اساس نمونه‌های کوچک بوده است. آن‌ها، گزارش‌های سالیانه را به‌طور الکترونیکی تحلیل نموده‌اند، اما تم‌های شناسایی شده از تحلیل عاملی را به‌طور دستی و ذهنی، تفسیر نموده‌اند. در نتیجه، آبراهامسون و پارک (۱۹۹۴) با تحلیل کامپیوتری نامه‌های بیش از ۱۰۰۰ مدیر، این موضوع را بررسی نمودند که چگونه مدیران شرکت، نتایج منفی را از سهامداران مخفی می‌کنند. برنامه کامپیوتری آن‌ها، فراوانی لغات را محاسبه می‌کند، درحالی‌که کدگذاران انسانی، کلمات دلالت‌کننده بر نتیجه منفی سازمانی را شناسایی می‌کنند. آبراهامسون و امیر (۱۹۹۶) از برنامه‌های کامپیوتری برای شناسایی پرکاربردترین لغات با تعابیر منفی در ۱۳۵۵ نامه برای سهامداران از ۱۹۸۸، استفاده نمودند، اما بازهم محتوای این لغات را به‌طور دستی تعیین کردند.

در تحقیقی مشابه، کلاتورثی و جونز (۲۰۰۳) اقدامات گزارش‌شده (در اظهارات مدیران که معادل بریتانیایی صحبت‌های مدیران است) در بیش از ۵۰ شرکت انتفاعی و ۵۰ شرکت غیرانتفاعی در انگلیس را با استفاده از روش‌های کامپیوتری برای استخراج داده‌هایشان، بررسی نموده و سپس از روش‌های دستی برای تفسیر مفاهیم استفاده کردند. آن‌ها نتیجه گرفتند که مدیران ش، روی خبرهای خوب تمرکز می‌کنند. آن‌ها همچنین دریافتند که مدیران در شرکت‌های در حال نزول، درباره علل زیربنایی عملکرد ضعیف خود بحث نمی‌کنند. همسو با

دیگر مطالعات و تئوری اسناد، آن‌ها دریافتند که مدیران، برای خبرهای خوب ارزش قائل‌اند و خبرهای بد درباره عوامل محیطی را سرزنش می‌کنند. با استفاده از همان نمونه از شرکت‌ها و یک متدولوژی مشابه، کلاتورثی و جونز (۲۰۰۶) مجدداً اظهارات روسا را تحلیل نموده و دریافتند که شرکت‌های سودآور در مقایسه با شرکت‌های غیر سودآور، کمتر بر شاخص‌های مالی کلیدی تمرکز نموده، از داده‌های کمی، مراجع شخصی کمتر، جملات مجهول استفاده نموده و بیشتر بر آینده تمرکز می‌کنند، بنابراین، نتایج کار قبلی مدیریت ادراکات را تأیید می‌کند.

تحلیل محتوای کامپیوتری

محققان حسابداری، تحلیل محتوای کامپیوتری از سال ۲۰۰۰ برای طیف وسیعی از متون شامل بخش‌های مختلف گزارش سالانه، گزارش‌های مطبوعات و مقالات خبری، بکار گرفته‌اند. اسمیت و تافلر (۲۰۰۰)، نشان می‌دهند که کلمات و مضامین کلیدی (مانند سود، سود سهام، گردش مالی و بهره‌وری) در گزارش‌های افشا اختیاری از اظهارات مدیران، شامل اطلاعات مهمی بوده و به‌طور نزدیکی مرتبط با عملکرد مالی است. با استفاده از تطبیق (OCP) و نرم‌افزار آماری (SPSS-X)، آن‌ها متغیرهای کلمات کلیدی و متغیرهای زمینه را بر اساس فراوانی وقوع آن‌ها محاسبه نموده و دریافتند که این متغیرها، به‌طور نظام‌مندی مرتبط با شکست یا موفقیت آینده شرکت است.

بک و همکاران (۲۰۰۱) و کلاچکنو و همکاران (۲۰۰۴)، قبل از پردازش بیشتر، کلمات را به مقادیر عددی تبدیل نموده‌اند. بک و دیگران (۲۰۰۱)، کدگذاری هوشمند را برای تبدیل کلمات به مقادیر عددی معرفی نموده و نقشه‌های خودسازمان‌دهی و نمودارهای هیستوگرام اسناد را برای تحلیل داده‌های کمی مالی و عملیات و داده‌های کیفی گزارش‌های مدیرعامل، بکار گرفته‌اند. رویکرد آن‌ها مبتنی بر کامپیوتر بوده و می‌تواند برای مقایسه اسناد متنی و تکمیل وظیفه بازیابی متن موتورهای جست‌وجو، استفاده شود. کلاچکنو و همکاران (۲۰۰۴)، از داده‌های متن‌کاوی و روش‌های متن‌کاوی برای بازیابی شاخص‌های عملکرد مالی آینده از بخش‌های کمی و کیفی گزارش‌های مالی، استفاده نموده‌اند. آن‌ها رویکرد کامپیوتری را اتخاذ نموده‌اند که هر واژه را به یک عدد، با توجه به طول و ارزش ASCII هر کاراکتر در کلمه، برمی‌گرداند. هیستوگرام‌ها، برای کلمات و جملات جمع‌آوری می‌شود. فواصل اقلیدسی بین هیستوگرام‌های

کلمات و هیستوگرام‌های جملات برای نمایش شباهت در واژگان و روش نگارش، اندازه‌گیری می‌شود.

گردس (۲۰۰۳)، یک تحلیل‌گر ادگار^۱ را بر اساس جست‌وجوی کلمات کلیدی توسعه داده و آن را ابزار مؤثری برای بازیابی بلوک‌های متن مرتبط با افشای Y2K از ۱۸۵۹۵ طرح 10-K، بین ۱۹۹۷ و ۱۹۹۹، بیان می‌داند. هاساینی و همکاران (۲۰۰۳) با به‌کارگیری یک بسته تحلیل متنی استاندارد برای جست‌وجوی متن گزارش‌های سالیانه الکترونیک، یک متدولوژی کامپیوتری برای محاسبه اتوماسیونی امتیازات افشا برای نمونه‌های بزرگ، ایجاد نموده‌اند. امتیازات افشای آن‌ها بر اساس جست‌وجوی کلمه، قابل‌مقایسه با آن‌هایی که توسط انجمن فدراسیون تحلیلگران مالی تحقیقات مدیریت (AIMR-FAF) توسعه یافته، است. بینون و دیگران (۲۰۰۵)، از یک تکنیک داده‌کاوی شناخته‌شده با عنوان تئوری مجموعه سخت به‌طور خاص مدل مجموعه‌های سخت متغیر به لحاظ دقت، برای پیش‌بینی موفقیت‌آمیز سودآوری شرکت بر اساس ویژگی‌های متنی (مثلاً مجموع لغات در اظهارات، درصد لغات مرتبط باخبر خوب، درصد لغات مرتبط باخبرهای بد) اظهارات مدیران، استفاده نموده‌اند.

ماگناسون و همکاران (۲۰۰۵)، نمایش داده‌های کمی از طریق نقشه‌های خودسازمان‌دهی شده را با نمایش داده‌های کیفی، با استفاده از شبکه‌های ترتیبی برای بررسی گزارش‌های دوره‌ای در صنایع مرتبط، ترکیب نموده‌اند. آن‌ها به این نتیجه رسیدند که تغییرات در داده‌های متنی، معمولاً موجب پیشبرد تغییرات در عملکرد مالی، در حداقل یک دوره مالی می‌شود. لی (۲۰۰۶)، فراوانی واژه‌های مرتبط با ریسک و عدم اطمینان را در ۳۴۱۸۰ طرح 10-K بین ۱۹۹۴ و ۲۰۰۵، به‌طور الکترونیکی شمارش نموده و آن را به‌عنوان شاخصی از تمایلات ریسک‌پذیری، استفاده نموده است. او دریافت که یک ارتباط آماری، بین افزایش در تمایلات ریسک‌پذیری و درآمدهای کمتر در آینده وجود دارد.

هنری (۲۰۰۶) با استفاده از تحلیل محتوای کامپیوتری یک الگوریتم داده‌کاوی درباره درآمدهای منتشرشده توسط گزارش‌های مطبوعات متغیرهای پیش‌بینی‌کننده شفاهی را شامل کلمات کلیدی حساب‌های مانند سود سهام، فروش و دیگر معیارها جهت دستیابی به تفاسیر محتوایی و سبکی عملکرد شرکت‌ها توسعه داد. این متغیرها و معیارها، موجب ارتقای صحت پیش‌بینی پاسخ بازار، تشخیص داده شدند. تتلاک و همکاران (۲۰۰۸) زبان مورد استفاده در

گزارش‌های خبری مالی را کمی نموده‌اند. مهم‌ترین یافته آن‌ها این بود که کلمات منفی در پیش‌بینی‌های خبری مالی، موجب کاهش درآمدهای شرکت می‌شود. هنری (۲۰۰۸)، تحلیل محتوای مبتنی بر کامپیوتر را از ۱۳۶۶ گزارش درآمد منتشر شده توسط مطبوعات بین ۱۹۹۸ و ۲۰۰۲ انجام داده و دریافت که لحن، بر واکنش‌های سرمایه‌گذاران تأثیر می‌گذارد. دمرز و وگا (۲۰۰۸) از یک برنامه تحلیل متنی برای اندازه‌گیری خوش‌بینی مدیریتی مطرح شده در متن گزارش‌های مطبوعاتی درآمدها استفاده کرد. بر اساس ۲۰۰۰۰ گزارش درآمد، آن‌ها شواهدی پیدا کردند که چنین خوش‌بینی را به بازگشت درآمد مرتبط می‌کرد. انگلبرگ (۲۰۰۸)، یک ابزار تحلیل وابستگی تاییپی، از پردازش زبان طبیعی برای محاسبه لغات منفی (طبق تعریف دیکشنری روانشناسی هاروارد)، در گزارش‌های درآمدها، بکار گرفته است. او دریافت که اطلاعات کیفی درآمد، قابلیت پیش‌بینی بیشتری برای قیمت‌داری‌ها، فراتر از اطلاعات کمی دارند.

دیویس و همکاران (۲۰۰۸)، از نرم‌افزار تحلیل متنی و محاسبه کلمات برای اندازه‌گیری سطح زبان خوش‌بینانه و بدبینانه در گزارش‌های درآمدهای مطبوعات، استفاده نمودند. این نوع استفاده از زبان، مرتبط با عملکرد آینده شرکت تشخیص داده شد. صدیق و همکاران (۲۰۰۸) از یک تحلیلگر محتوای کامپیوتری برای تحلیل آهنگ مقالات خبری و گزارش‌های درآمدی مطبوعات استفاده کردند. نتایج آن‌ها نشان می‌دهد که آهنگ مثبت، مرتبط با افزایش در بازگشت سهام شرکت و کاهش نوسانات می‌شود، در حالی که آهنگ منفی، مرتبط با کاهش در بازده‌ها و افزایش در نوسانات است.

اخیراً، مطالعات تحلیل محتوا، بر استنباط از موضوع تقلب تمرکز نموده‌اند. کچینی (۲۰۰۵) از ماشین‌های پشتیبانی برداری برای بررسی ریسک رخدادهای مالی مانند تقلب مدیریت، ورشکستگی و بیان دیگرگونه استفاده کردند. مدل بردار ورودی کشف تقلب، شامل هردوی متغیرهای مالی کمی (نسبت‌های حساب) و متغیرهای متنی کیفی (بردارهای مفهوم که با استفاده از آنتولوژی حسابداری پیش‌پردازش شده‌اند) برای شرکت‌های جعلی و غیر جعلی می‌شود. گول (۲۰۰۸) یک متدولوژی را برای کشف فعالانه تقلب با بررسی محتوای متنی کیفی گزارش‌های سالیانه با استفاده از ابزارهای پردازش زبانی طبیعی، ابداع نمودند. این متدولوژی، فن‌های یادگیری ماشینی را برای ایجاد یک طبقه‌بندی کننده خودکار بکار می‌گیرد که می‌تواند

احتمال تقلب را پیش‌بینی کند. در این مطالعه، یک تحلیل زبانی برای کشف ویژگی‌های زبانی (مانند صدا (فعال در مقابل منفعل)، نشانگرهای عدم اطمینان، شاخص خوانایی، آهنگ، استفاده از اسامی مناسب؛ نرخ‌های تایپ توکن (تغییرپذیری زبان) و غیره) انجام شده است که می‌تواند گزارش‌های سالیانه تقلبی را از گزارش‌های غیر تقلبی متمایز کند. نتایج نشان می‌دهد که تفاوت‌های نظام‌مندی در انتقال و شیوه نگارش گزارش‌های سالیانه تقلبی وجود دارد. صحت کشف تقلب، می‌تواند با افزودن یک بعد زبانی به بررسی تقلب، بسیار بهبود یابد. تحقیقات در تحلیل محتوای کامپیوتری نشان می‌دهد که روش‌های خودکار می‌توانند به‌طور موفقیت‌آمیزی جهت دستیابی و تحلیل نمونه‌های بزرگ استفاده شود و نتایج آن می‌تواند تعمیم‌پذیری بیشتری داشته باشد.

تحقیقات خوانایی در حسابداری

شاخه دیگری از تحقیق در گزارش‌های حسابداری، مطالعه خوانایی است. مطالعات خوانایی در حسابداری، تاریخچه طولانی دارد. ریاحی-بلکویی (۱۹۹۵) نتایج ۴۲ تحقیق انجام شده از ۱۹۵۲ تا ۱۹۹۳ را درباره خوانایی اسناد حسابداری، بخصوص اظهارات مالی و پانوشته‌های بیانیه‌های مالی، همراه با هشت تحقیق متمرکز بر قابلیت درک شرح دادند. از این مطالعات، نتیجه گرفتند که خوانایی و قابلیت درک به‌وسیله عواملان حرفه‌ای معتبر (مانند حسابداران عمومی معتبر)، غیر عواملان حرفه‌ای معتبر (مانند تحلیل‌گران مالی خبره) و افراد با تحصیلات کالج، کاربران غیر حرفه‌ای، مسئله‌دار است.

مطالعات اولیه، در اصل بر اساس کدگذاری دستی بود. کلاتورثی و جونز (۲۰۰۱) مطالعات خوانایی قبلی درباره گزارش‌های سالیانه را بررسی نمودند و آن‌ها را دارای نمونه‌های کوچکی از ۱۸ تا ۱۲۰ سند، تشخیص دادند. کدگذاری دستی هزینه‌بر نسخه‌های سخت، ظاهراً مانع از استفاده از نمونه‌های بزرگ می‌شود. کرتیس (۱۹۹۸) و کلاتورثی و جونز (۲۰۰۱)، مجموعه داده‌های خود را با استفاده از نرم‌افزار بسته فلش مایکروسافت ورد، برای محاسبه نمره فلش خوانایی گزارش‌های سالیانه الکترونیک، تکمیل نمودند.

مطالعات قبلی نشان‌دهنده تغییرات در خوانایی بخش‌های چندگانه همان گزارش‌های سالیانه و همچنین در همان بخش گزارش‌های سالیانه چندگانه است. برای مثال، خواندن پانوشته‌ها خیلی مشکل‌تر از اظهارات مدیران، تشخیص داده شد. اشکرودر و گیسون (۱۹۹۰) خوانایی

گزارش افشاشده ۴۰ شرکت را بررسی نموده و استنباط نمودند که خوانش MD&A و پانوشته‌ها بسیار سخت‌تر از نامه‌های مدیران است. کرتیس (۱۹۹۸) دریافت که اولین پاراگراف گزارش‌های سالیانه، آسان‌ترین بخش برای خوانش و پاراگراف دوم، سخت‌ترین است و به مدیران پیشنهاد نمود که اطلاعات نامطلوب را در پاراگراف دوم جای دهند.

کارهای اخیر درباره خوانایی، از ابزارهای زبان‌شناسی و پردازش کامپیوتری بهره گرفته‌اند. لی (۲۰۰۸) از یک بسته نرم‌افزاری پرل برای تعیین خوانایی گزارش‌های سالیانه استفاده نمود و خوانایی را مرتبط با درآمدها تشخیص داد. به‌طور خاص، گزارش‌های سالیانه شرکت‌هایی با درآمدهای پایین‌تر، کمتر خوانا هستند و شرکت‌هایی با گزارش‌های سالیانه خوانا، دارای درآمدهای مثبت پایدارتری هستند. گول و گانگولی (۲۰۰۹)، از نرم‌افزار GNU سبک، برای بررسی خوانایی گزارش‌های سالیانه شرکت‌های متقلب و غیر متقلب استفاده کردند. آن‌ها نشان دادند که شاخص خوانایی می‌تواند به‌عنوان یکی از ویژگی‌های متمایزکننده گزارش‌های سالیانه تقلبی از غیر تقلبی استفاده شود.

بازیابی اطلاعات حسابداری

بازیابی متن: واژه‌نامه حسابداری و نقش لغتنامه‌ها

بازیابی متن شامل مقایسه درخواست کاربر برای مجموعه‌ای از سوابق نوشتار آزاد به‌منظور انتخاب موارد مشابه مناسب است. یک عامل کلیدی در بازیابی موفقیت‌آمیز، انتخاب واژگان در فرموله سازی درخواست است. در علوم کتابخانه‌ای، لغتنامه‌ای که تغییرپذیری در زبان را لحاظ می‌کند، برای بهبود بازیابی استفاده می‌شود. فیدل (۱۹۹۱)، رفتار جست‌وجوی آنلاین و انتخاب توصیف‌گرها یا لغات متن را به‌عنوان کلیدهای جست‌وجو برای جست‌وجوی متن آزاد، بررسی نمود. او دریافت که هنگام استفاده از لغات متن، کلمات هم‌معنی برای جبران فقدان کنترل واژگان، استفاده نمی‌شوند. او، نیاز به دسترسی آنلاین آسان، انعطاف‌پذیر و ارزان را برای واژه‌نامه و ساخت زبان‌های قابل‌تبدیل کارکردی و سیستم‌های خبره واسطه‌ای مطرح می‌کند. مودامال (۱۹۹۸) اظهار داشت که در حوزه‌هایی که واژگان، اغلب تغییر می‌کند، یک اصطلاح‌نامه، نیاز به بروز رسانی دائمی دارد. برای مثال، اصطلاح‌نامه هنر و معماری آنلاین، به‌طور ماهیانه بروز می‌شود (اصطلاح‌نامه هنر و معماری آنلاین، ۲۰۱۰). جاکوب (۱۹۹۴)

دریافت که رشته‌های متفاوت، واژگان متفاوتی را برای توصیف یک پدیده استفاده می‌کنند که برای مجزا سازی بدنه دانش استفاده شده و از ارتباطات بین رشته‌ای، شامل «شناخت اشتراکاتی که بین محدوده‌های رشته‌ها وجود دارد» جلوگیری می‌کند. او از توسعه ساختارهای عمومی تری که انتقال اطلاعات بین حوزه‌ها را تسهیل نموده و گفتمان بین آن‌ها را تسهیل می‌کند، حمایت می‌کند.

تکنیک‌های ساخت اصطلاح‌نامه، معمولاً به رویکردهای دستی، خودکار، یا نیمه‌خودکار تقسیم می‌شود. ساخت دستی، پیچیده، هزینه‌بر و زمان‌بر است. نتیجه مستقیم آن این است که نگهداری (به‌روزرسانی) دستی، بندرت می‌تواند انجام گیرد. گانگولی و تام (۲۰۰۲) درباره اتخاذ روش‌ها و تکنیک‌های زبان‌شناسی کامپیوتری برای نسل خودکار دیکشنری‌های و اصطلاح‌نامه‌های حسابداری از مجموعه‌های موجود، حمایت می‌کنند. برای نشان دادن این رویکرد خودکار، آن‌ها واژه‌های با فراوانی بالا را از ۲۰۰۰ طرح ادگار با استفاده از جعبه‌ابزار زبان طبیعی CMU کمبریج، استخراج نمودند. رویکردهای دستی، به دلیل ناسازگاری‌های مطرح شده هنگام درگیر نمودن چندین نفر در این کار، موردانتقاد قرار گرفته‌اند. رویکردهای خودکار، از طرف دیگر، درحالی که به لحاظ زمان و هزینه، بر رویکردهای دستی برتری دارند، به دلیل ایجاد اصطلاح‌نامه‌هایی که به لحاظ ساختاری و محتوا ضعیف هستند و به دلیل ناتوانی‌شان در شناخت پیچیدگی‌های رابطه‌ای و ابهامات معنایی درونی در زبان‌ها، موردانتقاد هستند؛ بنابراین، افراد دیگر، اظهار می‌دارند که رویکردهای دستی و خودکار (به‌صورت یک رویکرد ترکیبی) مکمل هم بوده و ترکیب آن‌ها، سودمند و لازم است. گانگولی و وو (۲۰۰۰)، تحقیقات اولیه‌ای را با استفاده از معیارهای بررسی اسناد برای اکتشاف دسته‌بندی‌های خودکار مفاهیم حسابداری، انجام دادند. گارنسی (۲۰۰۶)، از شاخص معنایی مکنون (LSI) همراه با خوشه‌بندی تراکمی سلسله‌مراتبی برای گروه‌بندی خودکار واژگان مرتبط باهم و برای طبقه‌بندی موفقیت‌آمیز خودکار مفاهیم حسابداری، استفاده کردند. LSI، برای قرار دادن واژه‌ها در مضامین نزدیک به یکدیگر در یک فضای بررسی اسناد نسبت به استفاده از یک ماتریس بررسی سند بیشتر استفاده می‌شود. فیشر و گارنسی (۲۰۰۶) از LSI همراه با خوشه‌بندی تراکمی برای شناسایی اصلاحات FAS‌های مرتبط و برای تأیید طبقه‌بندی تغییرات در FAS‌ها استفاده

کرده‌اند. پژوهش‌های مقدماتی انجام‌شده به‌وسیله گارنسی (۲۰۰۸) تفاوت‌های فراوانی در واژه‌شناسی را نشان داد.

بروز نگه‌داشتن اصطلاح‌نامه‌ها، نیازمند شناسایی زود هنگام و مؤثر واژگان مربوطه است. گارنسی و فیشر (۲۰۰۸)، از تحلیل زبانی کامپیوتری در اظهارات معتبر و صورت‌های مالی برای شناسایی واژه‌های جدید حسابداری در کارشان روی این موضوع که چگونه واژه‌ها، در واژه‌نامه‌های حسابداری جای می‌گیرند، استفاده کردند. آن‌ها دریافتند که درحالی‌که واژگان حسابداری ابتدا در صورت‌های مالی ظاهر می‌شوند، تعداد زیادی از واژه‌های جدید، در ابتدا در ادبیات مالی ظاهر می‌شوند.

فیشر و گارنسی (۲۰۱۰) یک رویکرد ترکیبی را برای ساخت اصطلاح‌نامه در حسابداری بکار گرفتند. آن‌ها از اصطلاح‌نامه پرایس واتر هاوس ۱۹۷۴، به‌عنوان نقطه شروعی برای تحلیل ساختار سلسله‌مراتبی (درخت) واژه‌شناسی مرتبط با مزایای کارمندان که در ۱۹۷۴ موجود بود استفاده نمودند و سپس از روش‌های خودکار برای بررسی ادبیات حسابداری معتبری که امروزه موجود است (مانند تدوین اصول عمومی پذیرفته شده حسابداری بخش شماره ۷۱۵) به‌منظور تعیین اینکه چه واژگان جدید باید به درخت اضافه شود، استفاده کردند. نتیجه نهایی، یک اصطلاح‌نامه حسابداری دیجیتال برای زیرمجموعه‌ای از واژگان حسابداری است که می‌تواند برای تکمیل مکانیسم‌های جست‌وجوی مجموعه‌های آنلاین اسناد حسابداری، با ارائه واژگان فوری و اضافی مرتبط و بدین‌وسیله دستیابی به اطلاعات بازیابی شده از چنین مجموعه‌هایی، استفاده شود.

بازیابی فرمول‌بندی بررسی متن

بورتیک (۲۰۰۱) اثر ابهام در توسعه تحقیق را در پایگاه‌های داده رابطه‌ای بررسی نمود. آن‌ها، درخواست‌های اطلاعات را جمع‌آوری نموده و فواصل معنایی مختلف بین درخواست‌های اطلاعات و اظهار آن در زبان تحقیق که موجب ایجاد اطلاعات مطلوب می‌شود را تغییر دادند. در تحقیق آن‌ها، فاصله معنایی به‌عنوان «فاصله بین اطلاعات مورد تقاضای کاربران و بیان آن به زبان رابط تفحص که موجب ایجاد اطلاعات می‌شود»، تعریف شده است. آن‌ها دریافتند که صحت و کارایی رابطه معکوسی با ابهام در تقاضاهای اطلاعات دارند.

گارنسی و هوتالینگ (۲۰۰۷)، دریافتند که هنگامی که به یک گروه از آموزگاران حسابداری، فهرستی از واژگان حسابداری منتخب به صورت تصادفی از ادبیات حسابداری معتبر، داده می‌شود، حداقل یکی از آموزگاران حسابداری، نمی‌تواند هیچ واژه مرتبطی را برای ۱۳ واژه از ۳۰ واژه، شناسایی کند. آن‌ها دریافتند که تعداد واژگان بدون پاسخ، هنگامی که از آموزگاران خواسته می‌شود که واژه‌های مرتبط را از یک فهرست شناسایی کنند، کاهش می‌یابد. این درباره تحقیقات روانشناسی نیز صادق است که نشان داده‌اند که عموماً درخواست و تقویت نقش «برانگیزاننده‌ها» در تسهیل فرموله سازی تحقیق، بهتر است.

جامعه حسابداران رسمی آمریکا یکی از مهارت‌های مورد نیاز حسابداران سطح مقدماتی را توانایی در «انجام تحقیق و تحلیل اطلاعات» تعیین نموده است. گارنسی و همکاران (۲۰۰۹)، سودمندی تهیه فهرستی از تحقیقات انجام شده در ادبیات حسابداری معتبر را نشان دادند. واژگان پیشنهادی آن‌ها، به طور مستقیم از پایگاه داده FARS اقتباس شد و برای دانشجویان، ابزاری برای کاهش فاصله معنایی نیازهای اطلاعاتی‌شان و درخواست پایگاه داده توسط آن‌ها ایجاد نمود. وقتی به دانشجویان، فهرستی از واژگان جست‌وجوی ممکن داده می‌شود، توانایی آن‌ها برای یافتن توضیحات صحیح، به طور چشمگیری افزایش می‌یابد. به طور عام‌تر، تحقیقات در بازیابی اطلاعات، نشان می‌دهد که افراد نرخ موفقیت بیشتری در یافتن اهداف اطلاعاتی مرتبط به شکل واژگان شاخص مورد استفاده برای هر موضوع اضافه شده دارند.

بازیابی مقادیر حسابداری از متن

پایگاه داده ادگار یک منبع کلیدی عمومی از گزارش‌های مالی الکترونیک است. از آنجا که این گزارشات حسابداری نوعاً اسناد نیمه ساختارمند یا ساختار نیافته هستند، اختلاف زیادی در ساختار و واژگان فایل به چشم می‌خورد. استخراج اطلاعات مفید از این گزارش‌ها یک چالش است، بخصوص برای محققان دانشگاهی و تحلیل‌گران حرفه‌ای که قصد بررسی تعداد زیادی از شرکت‌ها را دارند. برخی محققان روی ایجاد نرم‌افزار هوشمند برای پشتیبانی گزارش‌های غیرساختارمند کار کرده‌اند. با افزودن به مطالعات پیشین، محققان حسابداری اخیراً پیشرفت‌های جدیدی در این حوزه داشته‌اند. لینمان و دگران (۲۰۰۱) رویه‌های متن‌کاوی را برای استخراج داده‌ها از صورت‌های مالی ارائه شده به شکل نگاره ASCII در بایگانی ادگار ارائه نموده‌اند. بو دیگران (۲۰۰۵)، بر دستیابی خودکار به اطلاعات مالی گنجانده شده در اسناد متنی

غیرساختارمند مانند بایگانی ادگار کار کرده‌اند. آن‌ها نمونه‌ای از گزارش‌های حسابرسی مالی شرکت را با دانش خالص سیستم FRAANK توسعه داده‌اند که تجزیه هوشمندی برای استخراج ارقام حسابداری از صورت‌های مالی با تطبیق صورت‌های مالی با موارد مشابه در XBRL بکار می‌گیرد. گرنٹ و کونلون (۲۰۰۶) اطلاعات گزینه‌های ذخیره کارمندی را از افشای صورت‌های مالی در پایگاه داده ادگار استخراج نموده‌اند. سیستم استخراج از ادگار توسط گرنٹ و کونلون (۲۰۰۶) از جعبه‌ابزار CMU KWIC برای مدل‌سازی زبان آماری، گزارش‌های منظم و تکنیک‌های تطبیق الگو استفاده می‌کند و برای شناسایی عبارات کلمه‌ای صورت‌های استانداردهای حسابداری مالی SFAS_ No. 123 آموزش داده شده است.

دینگ و چن (۲۰۰۶)، یک سیستم داده کاوی را برای استخراج خودکار داده‌های پرداخت از اوراق قرضه و سازمان بورس اوراق بهادار آمریکا ابداع نموده‌اند. آن‌ها دریافتند که تکنولوژی بازیابی انفرادی، برای استخراج وظیفه، ناقص است و یافتن یک رویکرد ترکیبی را نوید دادند. رویکرد ترکیبی آن‌ها، تطبیق الگوها را با توضیحات آن‌ها، وجود پایگاه دانش برای طبقه‌بندی کلمات، بررسی دامنه داده‌ها و تحلیل مضمون لغتی را یکپارچه می‌کند. کونگ و دیگران (۲۰۰۷)، یک رویکرد نظام‌مند را برای بازیابی ساختار و محتوا از صورت‌های مالی ارائه و عملکرد امیدبخشی را با این رویکرد، گزارش نموده‌اند. این چارچوب، از اطلاعات مختلف متنی، معنایی، هندسی و ساختاری مانند موقعیت نسبی یک صورت مالی مشخص، عنوان و سرفصل، درصد کاراکترهای نامشخص و حدود ترازوی، بهره می‌برد.

در مجموع ادبیات درباره بازیابی کمیت‌های حسابداری از اسناد متنی، دارای داده‌های کاملاً پراکنده‌ای است، اگرچه هر کدام از مطالعات سابق، موفقیت نسبی را نشان داده‌اند. با وجود اینکه کار بیشتری برای بهبود قابلیت دسترسی الکترونیک کمیت‌های حسابداری برای متن لازم است، اخیراً، مطالعاتی نشان می‌دهد که تطبیق الگوی مبتنی بر تفسیر و برنامه‌ریزی نظام‌مند، ابزارهای نویدبخشی هستند.

پیشنهادات تحقیقات آتی

همچنان که تکنولوژی دستیابی به اطلاعات را تسهیل می‌کند، نقش تحلیل متون و بازیابی اطلاعات در حوزه حسابداری از یک کاربرد ساده (مانند شناخت واژه) به کاربردهای پیچیده

(مانند ساخت اصطلاح‌نامه)، توسعه می‌یابد. نرم‌افزارهای بیشتری با استفاده از تکنیک‌های پردازش زبانی برای بازیابی اطلاعات و طبقه‌بندی، به‌منظور یافتن، استخراج و سازمان‌دهی اطلاعات لازم است. در این قسمت تلاش می‌شود با توجه به تحقیقات صورت گرفته و لزوم انجام تحقیقات بیشتر در این زمینه در جهت افزایش قابلیت درک و بازیابی اطلاعات برای استفاده‌کنندگان اطلاعات مالی موضوعات پژوهشی آتی تحت ۴ عنوان کلی (۱) میزان موفقیت تدوین‌کنندگان استاندارد در دستیابی به اهداف مدنظر خود (۲) ساخت‌واژه‌نامه‌های حسابداری، (۳) شناسایی واژه‌های جدید واردشده به واژه‌نامه و (۴) استانداردسازی واژگان حسابداری بیان شوند. در ادامه به‌طور مختصری این حوزه‌های بلقوه پژوهشی را توضیح خواهیم داد.

میزان موفقیت تدوین‌کنندگان استاندارد در دستیابی به اهداف مدنظر خود

ضروری است که یک تحقیق، به چگونگی آرشیو نمودن نسخه‌های مختلف ادبیات حسابداری معتبر پردازد بطوریکه قادر به دسترسی و بازیابی مؤثر آن باشد. درحالی‌که تغییرات در ادبیات حسابداری معتبر، متأثر از اصلاح مستقیم در متن استاندارد تدوین‌شده است، چگونگی دستیابی به نسخه‌های گذشته‌ی ادبیات حسابداری معتبر و همچنین اینکه تا چه حد از جزئیات به لحاظ زمانی معتبر می‌باشد، نامشخص است. سؤالات مرتبط با کاربرد درست اصول حسابداری، اغلب سالیان بسیار بعد از تاریخ بکارگیری مطرح می‌شود. برای مثال، درحالی‌که تحقیقات اوراق بهادر و کمیسیون مبادله انرون در ۲۰۰۱ آغاز شد، روند قانونی در این مورد از سال ۲۰۰۶ هنوز ادامه دارد. رخدادهای قبل از آن تاریخ نیاز به بررسی و قانونی شدن دارند که مستلزم توانایی بررسی قابلیت اجرای ادبیات حسابداری معتبر به‌عنوان یکی از نقاط پیشین به لحاظ زمانی است؛ بنابراین، دسترسی به نسخه‌های سابق ادبیات حسابداری معتبر، یک مسئله مهم است (فیشر، ۲۰۰۷). از طریق آن می‌توان میزان موفقیت تدوین‌کنندگان استاندارد در دستیابی به اهدافشان را سنجید و تأثیر استانداردهای تدوین‌شده در آن زمان را بررسی نمود. مثلاً فیشر (۲۰۰۹) نشان داد که الحاق یک ساختار XML رسمی می‌تواند از توانمندی‌های جست‌وجوی بهبودیافته برای ادبیات حسابداری معتبر با افزودن جست‌وجوی کلمات کلیدی واردشده توسط کاربر، با واژگان مرتبط اضافی مشخص‌شده، حمایت کند. مرزبندی واضح‌تر معانی (فراتر از شمارش لغات ساده) بخش‌های ادبیات حسابداری معتبر، می‌تواند از طریق برجسب‌های نشانه‌گذاری که امکان

جست‌وجوی‌های مؤثرتر کاربر را فراهم می‌کنند، توصیف شود. تحقیقات بیشتر درباره نشانه‌گذاری مؤثر، نیز می‌تواند مفید باشد.

ساخت واژه‌نامه‌های حسابداری

تمرکز FASB در سراسر پروژه تدوین، روی استانداردسازی واژگان مورد استفاده و همچنین کاهش تکرار اطلاعات، این موضوع را که کاربر قادر به فرموله بندی یک تحقیق که شامل زبان مورد استفاده در پایگاه داده است، مهم‌تر می‌سازد. این امر ممکن است بخصوص برای کاربرانی که غیر مجرب هستند یا اطلاعات مورد نیاز آن‌ها خوب تعریف نشده، مسئله‌ساز باشد. بدون نقاط دسترسی کافی، کاربران قادر نخواهند بود که اطلاعات مطلوب خود را بیابند. ساخت اصطلاح‌نامه‌ها، می‌تواند همچنین از تحلیل‌های موضوعی گزارش‌های حسابداری برای شناسایی طبقات محتوا بهره‌بردار. کارهای آینده در این حوزه، نیاز به تمرکز بر استخراج مفاهیم و ارتباطات جدید دارد بطوریکه بتوان این‌ها را به طبقاتی که منحصرأ با توضیح آن‌ها متناسب است، تخصیص داد.

شناسایی واژه‌های جدید وارد شده به واژه‌نامه

اغلب افراد، معنی واژه‌های جدید را تا زمانی که «وارد اخبار» نشوند، درک نمی‌کنند. برای اینکه افراد، امروزه قادر به فهم اخبار حسابداری باشند، لازم است با واژگان جدیدی که به واژه‌نامه حسابداری وارد می‌شوند، آشنا باشند. تحقیقات بیشتری لازم است تا مشخص شود که واژگان حسابداری برای اولین بار، در کجا معرفی شدند: حسابداری مالی یا دیگر اظهارات معتبر، گزارش‌های مالی شرکت، یا مجلات مالی دوره‌ای.

گارنسی و فیشر (۲۰۰۸)، استنباط نمودند که حوزه مالی، ممکن است یک منبع جایگزین برای واژگان جدید باشد. بسیاری واژه‌های جدید در حال ورود به واژه‌نامه حسابداری، عبارات چند لغتی هستند. ترکیبات، واژه‌هایی هستند که باید در کنار هم باشند و در کنار هم معنی دار هستند. روش‌های آماری مختلف، برای تعیین این موضوع استفاده شده‌اند که کدام کلمات پیاپی در متن، ترکیب منظم‌تری دارند. گارنسی و فیشر (۲۰۰۸)، از ابزار مدل‌سازی زبان CMU کمبریج، برای شناسایی واژگانی که باهم رخ می‌دهند و فراهم نمودن احتمالات و فراوانی‌های این واحدهای متوالی استفاده کرده‌اند. «هرس نمودن» واحدهای متوالی، برای حذف واحدهای

نامعمول بکار گرفته شده است که با حذف دستی واحدهای غیر مرتبط به حسابداری همراه است. تحقیقات آینده باید به بررسی حذف دقیق تر واحدهای متوالی که کمکی به زبان حسابداری از مدل زبان نمی کنند، بپردازد. برش های مختلفی باید مورد بررسی قرار گیرد و به امکان پذیری خالص سازی فهرست واحدهای شناسایی شده در نقطه خاص برش باید توجه شود. در نهایت، تحقیقات باید امکان پذیری به کارگیری روش های خودکار برای توسعه چشمگیر اندازه پیکره نوشته های مورد استفاده توسط گارنسی و فیشر (۲۰۰۸) بررسی نمایند. پیکره گسترده، باید شامل توضیحات مالی و همچنین منابع بی شماری از اطلاعات حسابداری مانند نشریات دوره ای مالی، استانداردهای حسابداری بین المللی و بیانیه های استانداردهای حسابداری باشد.

استانداردسازی واژگان حسابداری

یکی از مهم ترین دلایل برای استانداردسازی زبان، این است که به زبان امکان می دهد که به عنوان مدلی برای هر فرد در جامعه به منظور ارتباط قابل استفاده باشد. منظور از استانداردسازی زبان آواشناسی، املا، دستور زبان و واژگان اشاره است. به منظور استانداردسازی زبان حسابداری، ما به املا و واژگان مراجعه می کنیم.

استانداردسازی واژگان حسابداری مسئله مشکلی است زیرا نه تنها واژگان زبان حسابداری زیاد است، بلکه هر واژه نیاز به چندین ویژگی به لحاظ معناشناسی درست و کافی دارد. این موضوع در تعمیم دهی چالش ایجاد می کند. بررسی این امر که چگونه کلمات جدید در حوزه حسابداری شکل گرفته اند، یعنی کلمات جدید از یک کلمه اصلی شکل گرفته اند می تواند مفید باشد، بنابراین آیا کلمات جدید (مشتق شده) به کلمه اصلی مرتبط هستند. برای مثال، اگر یک لغت جدید به واژه نامه حسابداری وارد می شود، از یک لغت مبنا، با استفاده از پسوند یا پیشوند شکل گرفته و معانی کلمه جدید و کلمه مبنا ارتباط نزدیکی دارند، اگر این طور است پس این یعنی کلمه جدید، هم به طور رسمی و هم به لحاظ معنایی مرتبط با کلمه مبنا است (مثلاً امنیت با افزودن پسوندی به معنای «شدن یا تبدیل شدن»، به امن سازی تبدیل می شود). این نوع از تحلیل، به قوانین فرموله سازی شکل لغت کمک نموده و در حفظ و توسعه اصطلاح نامه، در پاسخ به تغییرات در حوزه حسابداری، نقش ایفا می کند. مثلاً FASB به تدوین کنندگان استاندارد دستور داده که حداقل تغییر در کلمات را هنگام نگارش موضوعات ایجاد نمایند. دستیابی به استانداردسازی در واژه نامه حسابداری ممکن است غیرممکن باشد. تحقیقات علوم

اطلاعات و تعامل انسان-کامپیوتر، نشان داده است که افراد از واژه‌های مختلف برای توصیف مفهوم مبتنی بر پیشینه‌های خود استفاده می‌کنند. حسابداری، مفاهیم را از مالی و اقتصاد اقتباس می‌کند. هر رشته، واژگان خاص خود را دارد و این امر مسئله را پیچیده‌تر می‌سازد. بعلاوه، تحقیقات اولیه توسط گارنسی (۲۰۰۸) مثال‌های متعددی از اختلافات زبانی بین اصول عمومی پذیرفته‌شده حسابداری بین‌المللی و ایالات متحده را نشان می‌دهد. تفاوت‌ها شامل واژه‌نامه و املا است. مشکلات استانداردسازی، الزام تحقیقات توسعه واژگان و اصطلاح‌نامه‌های بروز برای حسابداری را روز بروز بیشتر می‌کند.

بحث و نتیجه‌گیری

اسناد و اطلاعات مالی ارائه‌شده تأثیر مهمی بر تصمیمات استفاده‌کنندگان اطلاعات مالی دارد. مروری بر تحقیقات صورت گرفته در زمینه تحلیل و بازیابی متون ارائه‌شده بیانگر آن است که این حوزه پژوهشی جای رشد زیادی دارد و مستلزم انجام تحقیقات بیشتر در این زمینه است. روش‌های جدید خودکار در تحلیل‌های متنی و بازیابی اطلاعات، تبدیل به ابزارهای مهمی در حسابداری، مالی و تحقیقات کسب‌وکار شده است. در تحقیقات حسابداری و مالی، چنین فضایی در ویژگی‌های زبانی، منابع اسناد و متغیرهای اقتصادی موردنظر وجود دارد. تحقیقات اخیر متمرکز بر مجموعه محدودی از ویژگی‌های زبان‌شناسی کامپیوتری (مانند خوانایی و لحن) است. دیگر ابعاد تحلیل زبانی، نیازمند مطالعات آتی شامل جنبه‌های معنایی، گرامر، نحو و واج‌شناسی است (بخصوص آن‌هایی که به وسیله معانی تحلیل زبانی کامپیوتری قابل تعیین است). شواهد تجربی کمی در ارتباط با مشکلات معنایی در مفاهیم ارتباطی خارجی حسابداری وجود دارد (هارید، ۱۹۷۲) همچنین طراحی معانی فنی برای اصطلاحاتی که بار معنایی متفاوت در کاربرد عمومی یا سایر کاربردهای خوددارند و نبود واژه‌های استاندارد شده در گزارشگری مالی از مشکلات ساختاری موجود در این حوزه هستند. تحقیقات آتی در این زمینه را می‌توان در ۴ حوزه کلی (۱) میزان موفقیت تدوین‌کنندگان استاندارد در دستیابی به اهداف مدنظر خود (۲) ساخت واژه‌نامه‌های حسابداری، (۳) شناسایی واژه‌های جدید واردشده به واژه‌نامه و (۴) استانداردسازی واژگان حسابداری بیان کرد. تاکنون تحقیقات در حوزه تحلیل متن و بازیابی اطلاعات و تأثیر آنها بر درک استفاده‌کنندگان اطلاعات مالی صورت نگرفته است بنابراین به

محققان پیشنهاد می‌شود با توجه به حوزه‌های پیشنهادی در این پژوهش به این موضوعات جالب و تأثیرگذار پژوهشی پردازند.

منابع

- Abrahamson, E. and E. Amir. 1996. The information content of the president's letter to shareholders. *Journal of Business Finance & Accounting* 23 _8_: 1157-1182.
- Back, B. J. Toivonen, H. Vanharanta, and A. Visa. 2001. Comparing numerical data and text information from annual reports using self-organizing maps. *International Journal of Accounting Information Systems* 2: 249-269.
- Belkaoui, A. 1980. The interprofessional linguistic communication of accounting concepts: An experiment in sociolinguistics. *Journal of Accounting Research* 18 _2_: 362-374.
- Beynon, M. J. M. A. Clatworthy, and M. J. Jones. 2005. The prediction of profitability using accounting narratives: A variable-precision rough set approach. *Intelligent Systems in Accounting and Finance Management* 12: 227-242.
- Boo, E. and R. Simnett. 2002. The information content of management's prospective comments in financially distressed companies: A note. *Abacus* 38 _2_: 280-295.
- Borthick, A. F. P. L. Bowen, D. R. Jones, and M. H. K. Tse. 2001. The effects of information request ambiguity and construct incongruence on query development. *Decision Support Systems* 32: 3-25.
- Brennan, N. M. E. Guillamon-Saorin, and A. Pierce. 2009. Impression management: Developing and illustrating a scheme of analysis for narrative disclosures—A methodological note. *Accounting, Auditing & Accountability Journal* 22 _5_: 789-832.
- Burke, J. F. 1997. Report on standards overload. *The CPA Journal* 66 _3_: 11.
- Clatworthy, M. and M. J. Jones and. 2003. Financial reporting of good news and bad news: Evidence from accounting narratives. *Accounting and Business Research* 33 ۱۸۵-۱۷۱ :۳.
- Clatworthy, M. and M. J. Jones and. 2006. Differential patterns of textual characteristics and company performance in the chairman's statement. *Accounting, Auditing & Accountability Journal* 19 ۵۱۱-۴۹۳ :۴.
- Clatworthy, M. and M. J. Jones. 2001. The effect of thematic structure on the variability of annual report readability. *Accounting, Auditing & Accountability Journal* 14 ۳۲۶-۳۱۱ :۳.
- Cong, Y. A. Kogan, and M. Vasarhelyi. 2007. Extraction of structure and content from the EDGAR database: A template-based approach. *Journal of Emerging Technologies in Accounting* 4 ۸۶-۶۹ :۱.

- Davis, A. J. Piger, and L. Sedor. 2008. Beyond the numbers: Managers' use of optimistic and pessimistic tone in earnings press releases. Working paper, University of Oregon.
- Demers, E. and C. Vega. 2008. Soft information in earnings announcements: News or noise? Working paper, Board of Governors of the Federal Reserve System.
- Ding, C. and P. Chen. 2006. Mining executive compensation data from SEC filings. In Proceedings of the 10th International Conference on Data Engineering Workshops, 1-7.
- Engelberg, J. 2008. Costly information processing: Evidence from earnings announcements. Working paper, Northwestern University.
- Fidel, R. 1991. Searchers' selection of search keys: II. Controlled vocabulary or free-text searching. *Journal of the American Society for Information Science* 42: 514-517.
- Fisher, I. E and M. R. Garnsey and. 2010. Improving information retrieval from accounting documents: A prototype digital thesaurus for employee benefits. Paper presented at the 2010 AAA Midyear Meeting of the Information Systems and the Strategic and Emerging Technologies Sections January, Clearwater, FL
- Fisher, I. E and M. R. Garnsey. 2006. The semantics of change as revealed through an examination of financial accounting standards amendments. *Journal of Emerging Technologies in Accounting* 3: 41-59.
- Fisher, I., Garnsey, M. R., Tam, K. (2010). The Role of Text Analytics and Information Retrieval in the Accounting Domain. *Journal of emerging technologies in accounting*. Vol. 7. pp. 1-24
- Gangolly, J. 1995. Some thoughts on the engineering of financial accounting standards. In *Artificial Intelligence in Accounting and Auditing III*, edited by M. A. Vasarhelyi, pp. 177-190. New York, NY: Markus Weiner Publishing.
- Garnsey, M. R. A. 2008 preliminary examination of differences in language between U. S. GAAP and international GAAP. *Collected Papers of the 17th Annual Research Workshop on Artificial Intelligence and Emerging Technologies in Accounting Auditing and Tax*, Anaheim, CA.
- Garnsey, M. R. A. 2009 comparison of language in the Codification and the FASB and its predecessors pronouncements, paper presented at the AAA Information Systems Section Midyear Meeting, Charleston, SC.
- Garnsey, M. R and A. W. Hotaling. 2007. Improving identification of search terms in the FARS database. *Review of Business Information Systems*. 11: 45-55.
- Garnsey, M. R and I. E. Fisher. 2008. Appearance of new terms in accounting language: A preliminary examination of accounting pronouncements and financial statements. *Journal of Emerging Technologies in Accounting* 5: 1-17.

- Garnsey, M. R. 2006. Automatic classification of financial accounting concepts. *Journal of Emerging Technologies in Accounting* 3 :۲۱-۳۹ :۱.
- Gerdes, J. 2003. EDGAR-analyzer: Automating the analysis of corporate data contained in the SEC's EDGAR database. *Decision Support Systems* 35 :۷-۲۹ :۱.
- Goel, S. 2008. Qualitative information in annual reports and the detection of corporate fraud: A natural language processing perspective. Doctoral dissertation, University at Albany-SUNY.
- Goldberg, L. 1965. *An Inquiry into the Nature of Accounting*. Sarasota, FL: AAA.
- Ingrid E. Fisher, Margaret R. Garnsey, Sunita Goel, and Kinsun Tam (2010) The Role of Text Analytics and Information Retrieval in the Accounting Domain. *Journal of Emerging Technologies in Accounting*: December 2010, Vol. 7, No. 1, pp. 1-24.
- Jacob, E. K. 1994. Classification and crossdisciplinary communication: Breaching the boundaries imposed by classificatory structure. In *Advances in Knowledge Organization, Vol.: Knowledge Organization and Quality Management*, edited by H. Albrechtsen, and S. Oernager, 101-108. Frankfurt/Main, Germany: Indeks Verlag.
- Jones, M. J , - and P. A. Shoemaker. 1994. Accounting narratives: A review of empirical studies of content and readability. *Journal of Accounting Literature* 14: 142-184.
- Kloptchenko, A. T. Eklund, J. Karlsson, B. Back, J. Vanharanta, and A. Visa. 2004. Combining data and text mining techniques for analyzing financial reports. *Intelligent Systems in Accounting and Finance Management* 12 :۲۹-۴۱ :۱.
- Leinnemann, C. F. Schlottmann, D. Seese, and T. Stuempert. 2001. Automatic extraction and analysis of The Role of Text Analytics and Information Retrieval 21 *Journal of Emerging Technologies in Accounting* Volume 7, 2010 American Accounting Association
- Li, F. 2006. Do stock market investors understand the risk sentiment of corporate annual reports? Working paper, University of Michigan.
- Miller, G. S. 2002. Earnings performance and discretionary disclosure. *Journal of Accounting Research* 40 :۱۷۳-۲۰۳ :۱.
- Muddamalle, M. R. 1998. Natural language versus controlled vocabulary in information retrieval: A case study in soil mechanics. *Journal of the American Society for Information Science* 49 :۸۸۷-۸۸۱ :۱۰.
- Schroeder, N. and C. Gibson. 1990. Readability of management's discussion and analysis. *Accounting Horizons* ۸۷-۷۸ :۴ ۴.
- Tetlock, C. M. Saar-Tsechansky, and S. Macskassy. 2008. More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance* 63: 1437-1467.