Constructing an Ontology of the Persian Stock Market and Financial Markets

Mohammad Hossein Samani
PhD Candidate in Information Technology; Faculty of Industries and Systems; Tarbiat Modares University; Tehran, Iran;
Email: mhsamani@modares.ac.ir

Amir Albadvi*
PhD in information systems; Professor; Faculty of Industries and Systems; Tarbiat Modares University; Tehran, Iran;
Email: albadvi@modares.ac.ir

Iranian Journal of Information Processing and Management

Received: 06, Nov. 2022 **Accepted: 25, Feb. 2023**

Iranian Research Institute for Information Science and Technology (IranDoc)
ISSN 2251-8223
eISSN 2251-8231
Indexed by SCOPUS, ISC, & LISTA
Vol. 39 | No. 1 | pp. 101-130
Autumn 2023
https://doi.org/jipm.39.1

Abstract: It is very difficult to predict stocks and commodity price index due to the presence of many and influential uncertainties. With the help of the accumulated information available in the current digital age and the power of high-performance computing machines, there is a lot of focus on designing algorithms that can learn stock market trends and successfully predict stock prices. Therefore, it will be very useful to create appropriate knowledge bases in order to increase the accuracy and efficiency of these systems and to facilitate the routine of using conventional knowledge in machine learning systems. The purpose of this research is to develop a Persian ontology for modeling the stock market and identifying factors affecting the stock market. The created ontology will lead to the enrichment and completion of the capacities of the existing knowledge bases in this field. For this purpose, in this research, a domain-specific ontology has been developed in the field of stock market and financial markets, which was prepared in Persian language by the authors of this research. After introducing this ontology, the details of the steps required to collect relevant data, semi-automated development and evaluation of this knowledge resource are described. The constructed ontology includes 565 concepts, 496 hierarchical relationships, 137 non-hierarchical relationships, and 937 samples that have been evaluated with various criteria and have a favorable status. It seems that this ontology in the current conditions and according to the evaluated volume and quality, is quite suitable to be used as a source of knowledge to improve the performance of machine learning systems for stock forecasting, and it can also be used to training stock market analysts and creating a knowledge base for brokerages, improving the process of retrieving semantic information and helping to determine the investment strategies of individuals in investment funds.

Keywords: Ontology, Ontology Learning, Stock Market, Financial Markets, Stock Forecasting

* Corresponding Author

ساخت هستان‌نگاری بورس و بازارهای مالی فارسی

محمدحسین ثمنی
دانشجوی دکتری سیستم‌های اطلاعاتی؛ دانشکده صنایع و سیستم‌ها؛ دانشگاه تربیت مدرس؛ تهران، ایران mhsamani@modares.ac.ir

امیر البدوی
دکتری فناوری اطلاعات؛ استاد؛ دانشکده صنایع و سیستم‌ها؛ دانشگاه تربیت مدرس؛ تهران، ایران؛ پدیدآور رابط albadvi@modares.ac.ir

پژوهش‌نامهٔ پردازش و مدیریت اطلاعات

دریافت: ۱۴۰۱/۰۸/۱۵ | پذیرش: ۱۴۰۱/۱۲/۰۶ | مقاله برای اصلاح به مدت ۱۵ روز نزد پدیدآوران بوده است.

نشریهٔ علمی | رتبهٔ بین‌المللی
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)
شاپا (چاپی) ۸۲۲۳-۲۲۵۱
شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در LISTA، ISC، و SCOPUS
jipm.irandoc.ac.ir
دورهٔ ۳۹ | شمارهٔ ۱ | ص‌ص ۱۰۱-۱۳۰
پاییز ۱۴۰۲
https://doi.org/jipm.39.1

چکیده: پیش‌بینی سهام و شاخص قیمت سهام و کالا به‌دلیل وجود عدم قطعیت‌های فراوان و تأثیرگذار بسیار دشوار است. با کمک اطلاعات انباشته موجود در عصر دیجیتال فعلی و قدرت ماشین‌های محاسباتی قوی، تمرکز زیادی بر طراحی الگوریتم‌هایی وجود دارد که می‌توانند روند بازار سهام را بیاموزند و قیمت‌های سهام را با موفقیت پیش‌بینی کنند. بنابراین، ایجاد پایگاه‌های دانش مناسب در راستای افزایش دقت و بهره‌وری این سیستم‌ها و تسهیل روال استفاده از دانش عرفی موجود در سیستم‌های یادگیری ماشینی بسیار مفید خواهد بود. هدف از این پژوهش توسعه هستان‌نگاری فارسی برای مدل‌سازی حوزه بورس و تشخیص عوامل تأثیرگذار بر بازار سهام است. هستان‌نگاری ایجادشده منجر به غنی‌سازی و تکمیل ظرفیت‌های پایگاه‌های دانش موجود در این حوزه خواهد شد. بدین منظور در این پژوهش یک هستان‌نگاری خاص دامنه در حوزه بورس و بازارهای مالی توسعه داده‌شده که به زبان فارسی توسط نویسندگان این پژوهش تهیه شده است. پس از معرفی این هستان‌نگاری، جزئیات گام‌های مورد نیاز برای جمع‌آوری دادگان مرتبط، توسعه نیمه‌خودکار و ارزیابی این منبع دانش بیان می‌گردد. هستان‌نگاری ساخته‌شده شامل ۵۶۵ مفهوم، ۴۹۶ رابطه سلسله‌مراتبی، ۱۳۷ رابطه غیرسلسله‌مراتبی و ۹۳۷ نمونه است که با معیارهای مختلفی ارزیابی شده و وضعیت مطلوبی دارد. به نظر می‌رسد این هستان‌نگاری در شرایط فعلی و با توجه به حجم و کیفیت ارزیابی‌شده، برای استفاده به‌عنوان منبع دانشی برای بهبود عملکرد سیستم‌های یادگیری ماشین برای پیش‌بینی سهام به‌طور کامل مناسب بوده و همچنین، می‌توان از آن در جهت آموزش تحلیلگران بورس و ایجاد پایگاه دانش برای

کارگزاری‌ها، بهبود فرایند بازیابی اطلاعات معنایی و کمک به تعیین استراتژی‌های سرمایه‌گذاری افراد در صندوق‌های سرمایه‌گذاری استفاده نمود.

کلیدواژه‌ها: هستان‌نگاری، یادگیری هستان‌نگاری، بورس، بازارهای مالی، پیش‌بینی سهام

۱. مقدمه

۱-۱. روش پژوهش

در این پژوهش از رویکرد استدلال ابداعی[1] استفاده می‌شود. این رویکرد که شکلی از استنتاج منطقی است، با مشاهده یا مجموعه‌ای از مشاهدات شروع می‌شود و سپس، به‌دنبال ساده‌ترین و محتمل‌ترین نتیجه‌گیری از مشاهدات است. این فرایند، بر خلاف استدلال استنتاجی، هرچند نتیجهٔ معقولی به‌دست می‌دهد، اما به‌طور قطعی آن را اثبات نمی‌کند. بنابراین، نتیجه‌های استدلال ابداعی دارای ردّی از عدم اطمینان یا تردید دانسته می‌شوند، و با اصطلاحاتی مانند «بهترین گزینهٔ موجود» یا «به احتمال زیاد» توصیف می‌شوند. کمابیش می‌توان استدلال ابداعی را به‌عنوان «استنتاج به بهترین توضیح» بیان نمود.

۱-۲. معرفی هستان‌نگاری

اگر بخواهیم تعریف اولیه‌ای از هستان‌نگاری (یا هستان‌شناسی)[2] داشته باشیم، باید گفت هستان‌نگاری به‌عنوان یک واژه، ترکیبی است از دو کلمه آنتو (هستی) و لوژی (مطالعه) که در ابتدا، در دانش فلسفه برای مطالعه وجود و هستیِ موجودیت‌های جهان از طریق دسته‌بندی جامعی از آن‌ها ظهور کرده است. هستان‌نگاری در علوم کامپیوتر و به‌ویژه زمینه معناشناسی، برای مدل‌سازی و نمایش مفهوم کلمات و ارتباط بین آن‌ها در یک زمینه دلخواه به‌کار گرفته می‌شود. به‌طور کلی، می‌توان گفت که هستان‌نگاری بیانی قابل فهم از مفاهیم دانشی یک زمینه دلخواه است که به‌منظور ایجاد امکان اشتراک و استفاده مجدد دانش ارائه شده است. این تعریف کلی، از تعاریف بیان‌شده در منابع علمی مختلف جمع‌بندی شده است که برخی از آن‌ها در ادامه آمده‌اند.

با توجه به اهمیت و نقش هستان‌نگاری در سیستم‌های مبتنی بر دانش، تعاریف

1. abductive reasoning
2. ontology

متنوعی از هستان‌نگاری در مراجع مختلف آورده شده است:

- نظریه‌ای برای موجودیت‌های ذهن یک عامل هوشمند (Schreiber, Wielinga, and Breuker 1993)؛
- درک مشترکی از برخی حوزه‌های دلخواه (Uschold and Gruninger 2004)؛
- توافق‌هایی در مورد تصورات مشترک ذهنی (Chandrasekaran, Josephson, and Richard 1999) که شامل یک چارچوب مفهومی برای مدل‌سازی دانش یک حوزه است؛
- یک طبقه‌بندی مفاهیم برای بدنه دانشی یک دامنه یا وظیفه مشخص برای تفسیر معنایی آن دانش (Chou, Vassar, and Binshan 2008)؛
- یک توصیف صریح و جزئی برای یک مفهوم به‌منظور ایجاد امکان طراحی ماژولار و مجدد بخش‌های یک سیستم دانش‌بنیان که قابل بیان در سطح بالاتر برای مجموعه‌ای از نظریه‌های محتمل روی حوزه مورد نظر است (Van Heijst et al. 1995)؛
- سلسله‌مراتبی از کلاس‌ها و مجموعه‌ای از روابط برای تعریف رده‌بندی آن‌ها و محدودیت‌هایی برای حفظ جامعیت آن‌ها (Gruber 1994). «گروبر» با این تعریف ادعا می‌کند که برای هوش مصنوعی مفاهیمی قابل استفاده هستند که قابل بازنمایی باشند. در نتیجهٔ این تعریف می‌توان گفت که تنها موجودیت‌هایی (همراه با صفات بازنمایی‌شده) قابل قبول هستند که در سیستم اطلاعاتی بازنمایی شده باشند. در این معنا، هستان‌نگاری چارچوبی است برای مفهوم‌سازی: دیدگاهی انتزاعی و ساده‌شده از جهان که ما به کمک آن، جهان را به‌منظور خاصی بازمی‌نماییم.
- با توجه به همه موارد بالا، با این فرض که هستان‌نگاری‌ها ابزارهای مهم معنایی هستند و در کاربرد هوش مصنوعی باید تصویری موضوعی از حوزهٔ مورد نظر با بهره‌گیری از مفاهیم و روابط میان آن‌ها ایجاد کنند، این تعریف را برای هستان‌نگاری درنظر می‌گیریم: «هستان‌نگاری، یک تعریف صریح و صوری از یک مفهوم‌سازی ذهنی اشتراکی است» (Gruber 1994).

با علم به این موضوع که این تعریف که یکی از مشهورترین تعاریف هستان‌نگاری است، به نظر ضروری می‌رسد که بخش‌های مختلف این تعریف شفاف‌سازی شود: مفهوم‌سازی، گام اول در طراحی یک هستان‌نگاری خواهد بود و لازم است در بخش طراحی، مفاهیم موجود در زمینه مورد نظر به همراه روابط معنایی میان آن‌ها و محدودیت‌های

مربوط به هر مفهوم بدون ابهام مشخص شود. واژه «صوری» بیانگر قابل فهم و دقیق بودن برای رایانه است. یعنی مفاهیم و روابط موجود در یک حوزه باید به شیوه‌ای قابل فهم و بدون ابهام برای رایانه بیان شوند. بنابراین، لازم است برای هستان‌نگاری‌ها زبان و مجموعه‌ای از قوانین مختص به خود تعریف گردد که برای رایانه نیز معنادار باشد تا بدین وسیله امکان استنتاج و کشف دانش ایجاد گردد. عبارت «اشتراکی» بیانگر این است که هستان‌نگاری باید بتواند امکان درهم‌کنش و تبادل دانش را بین ماشین‌ها و نظام‌های اطلاعاتی مختلف برقرار کند.

با توجه به این تعریف، هستان‌نگاری شامل یکسری مفهوم، رابطه، نمونه و ویژگی است و از طریق آن دو هدف متصور است:

۱. ارائه یک فهم مشترک از یک مفهوم واحد (برای رفع ابهام در مواردی چون بیان یک مفهوم با دو کلمه یا بیان یک کلمه برای دو مفهوم)؛
۲. ایجاد ارتباط میان مفاهیم در وب و دنیای واقعی برای درک ماشین. (توصیف رسمی دانش با مشخص‌کردن مفاهیم و موجودات یک دامنه و بیان روابط بین آن‌ها به‌صورت رسمی و قابل فهم توسط ماشین و استفاده از قواعد مشخص).

۳-۱. تمرکز و ساختار مقاله

گفتیم که طبق تعریف، هستان‌نگاری عبارت است از یک تعریف صریح و رسمی برای درک یک حوزهٔ خاص. ساختن هستان‌نگاری از فعالیت‌های زمان‌بر و دقیق بوده و به مشارکت افراد متخصص در دو زمینه مهندسی دانش و هستان‌نگاری و همچنین افراد خبره در حوزهٔ مورد نظر برای ساخت هستان‌نگاری نیاز است. ساخت هستان‌نگاری به‌طور عام به سه شیوه قابل انجام است:

- ◇ کاملاً دستی و توسط افراد خبره در حوزه ساخت هستان‌نگاری؛
- ◇ نیمه‌خودکار و توسط ابزارها و نرم‌افزارها و با دخالت انسان؛
- ◇ کاملاً خودکار.

به‌علت زمان‌بر بودن، هزینهٔ بالا و نداشتن انعطاف برای تغییرات محتمل، انتظار ساخت هستان‌نگاری به‌صورت دستی و توسط افراد خبره در آن حوزه معقول نیست. ساخت کاملاً خودکار هستان‌نگاری نیز در حالت کلی غیرممکن است، زیرا به دانش و دخالت افراد خبرهٔ حوزه برای افزودن اطلاعات برای ساخت هستان‌نگاری، تأکید روی بخش خاصی از اطلاعات و همچنین بهبود

و تغییر هستان‌نگاری ساخته‌شده نیاز است. در نتیجه، بهترین روش، توسعه نیمه‌خودکار هستان‌نگاری و با دخالت اندک انسان است. در این راستا به‌طور عام لازم است از اکتشاف دانش[1]، پردازش زبان طبیعی[2]، بازیابی اطلاعات[3]، یادگیری ماشین، استنتاج[4] و مدیریت پایگاه داده استفاده نمود.

یادگیری هستان‌نگاری در حقیقت شاخه‌ای از استخراج اطلاعات است که هدف آن ساخت (نیمه)خودکار هستان‌نگاری از مجموعه‌ای از اسناد (پیکره) یا دیگر پایگاه‌های داده است. یادگیری هستان‌نگاری می‌تواند از منابع داده‌ای مختلف مانند متن، فرهنگ لغت و یا پایگاه دانش باشد. از منظر استخراج دانش، هستان‌نگاری نوع دیگری از مدل است (و البته پیچیده‌تر از مدل‌های یادگیری ماشین) هستند که می‌بایست توسط یک زبان فرضی بیان شوند. به‌طور کلی، یادگیری هستان‌نگاری برقرارکردن یک نگاشت بین مؤلفه‌هاست که برخی موجودند و برخی موجود نیستند و هدف، یافتن اجزای مجهول است.

با این مقدمه در این پژوهش در راستای یادگیری هستان‌نگاری به‌صورت عملی گام‌های زیر انجام شده است: استخراج مفاهیم (ترم‌ها)، استخراج مفاهیم، استخراج سلسله‌مراتب مفاهیم و استخراج سایر روابط، ویژگی‌ها، قوانین و نمونه‌ها.

این مقاله به شرح زیر سازماندهی شده است: بخش ۲، کارهای مرتبط را بررسی می‌کند. بخش ۳، فرایند ساخت هستان‌نگاری فارسی بورس و بازارهای مالی را معرفی می‌کند و روش تولید آن را به تفصیل توضیح می‌دهد. در بخش ۴، جزئیات ارزیابی‌ها و بحث مقاله ارائه شده است. سرانجام، در بخش ۵، زمینه‌های کاربردی معرفی شده و نتیجه‌گیری صورت می‌گیرد.

۲. پژوهش‌های مرتبط

در این بخش، پژوهش‌های مرتبط را در دو دسته اصلی مرور می‌کنیم: ۱) پژوهش‌های حوزه ساخت هستان‌نگاری (برای معرفی روش‌های رایج توسعه نیمه‌خودکار هستان‌نگاری)، و ۲) روش‌های بیان‌کننده کاربرد هستان‌نگاری در بازارهای مالی (در جهت بیان ضرورت و روش‌های استفاده از هستان‌نگاری).

۲-۱. پژوهش‌های حوزه ساخت نیمه‌خودکار هستان‌نگاری

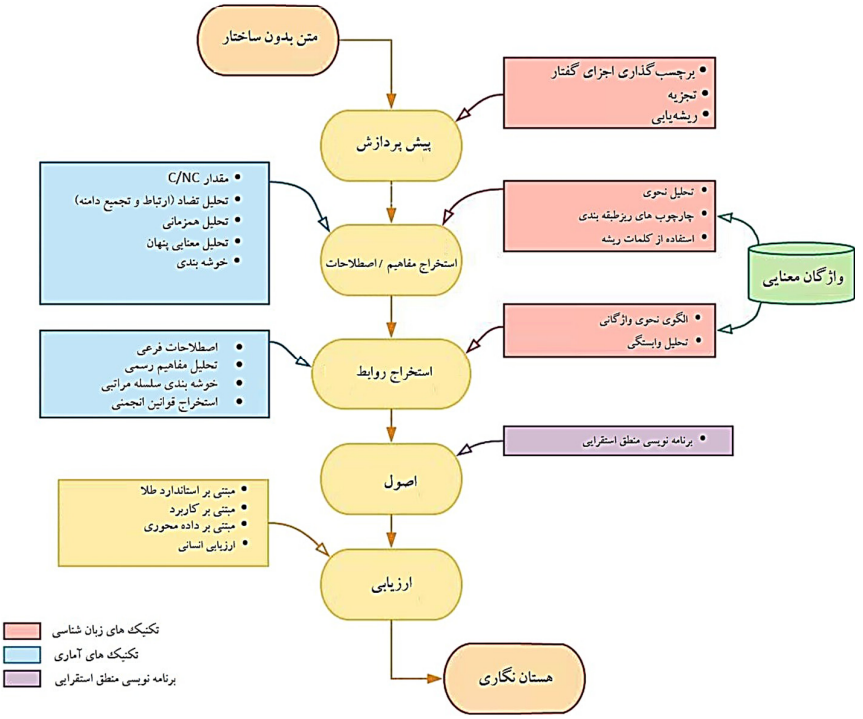
در زمینه ساخت نیمه‌خودکار هستان‌نگاری چند مطالعه مروری جامع وجود دارند. در

1. knowledge discovery
2. natural language processing
3. information retrieval
4. reasoning

سال ۲۰۰۲، پژوهش «دینگ و فو» ویژگی‌های ۱۲ سیستم یادگیری هستان‌نگاری را خلاصه کرد و جزئیاتی در مورد الگوریتم‌های مختلف یادگیری ارائه نمود. یافته‌های این بررسی به‌شرح زیر خلاصه می‌شود: (۱) بیشتر سیستم‌های یادگیری هستان‌نگاری به‌جای ساختن آن از ابتدا، هستان‌نگاری را با کمک مفاهیم اولیه یا هستان‌نگاری پایه یاد می‌گرفتند؛ (۲) تکنیک‌های پردازش زبان طبیعی برای استخراج مفهوم، نتایج امیدوارکننده‌ای را نشان داده، در حالی که استخراج رابطه یکی از چالش‌های اصلی در پردازش زبان طبیعی بود (Ding & Foo 2002). در سال ۲۰۰۳، کنسرسیوم OntoWeb گزارشی درباره ۳۶ روش و ابزار مرتبط مورد استفاده برای یادگیری هستان‌نگاری از متن بدون ساختار منتشر کرد.

در سال ۲۰۰۴، «شمس فرد و عبدالله‌زاده بارفروش» در یک مطالعه مروری سیستم‌های مختلف یادگیری هستان‌نگاری را بر اساس سه دسته زیر طبقه‌بندی و مقایسه کردند: (۱) نقطه شروع سیستم یادگیری هستی‌نگاری (یعنی هستی‌نگاری از ابتدا ساخته شده است یا از یک هستی‌نگاری پایه از پیش ساخته‌شده استفاده می‌کند)، (۲) چه نوع هستان‌نگاری برای یک کاربرد مورد نیاز است (مثلاً یک برنامه علمی ممکن است به یک هستان‌نگاری کوتاه و بدیهی برای حل مشکلات خود نیاز داشته باشد)، و (۳) درجه خودکار بودن (Shamsfard & Abdollahzadeh Barforoush 2004).

در ادامه، مطالعات مروری دیگری نیز منتشر شدند؛ مثل Buitelaar, Cimiano and Magnini (2005); Zhou (2007); Hazman, El-Beltagy and Rafea (2011); و Asim et al. (2018) که هر یک از دیدگاه خاصی روش‌های یادگیری هستان‌نگاری را مرور کرده است. جدیدترین پژوهش روش‌های ساخت خودکار هستان‌نگاری را به شکل فلوچارت زیر معرفی کرده است.



شکل ۱. انواع روش‌های یادگیری هستان‌نگاری (Asim et al. 2018)

در پژوهش Asim et al. (2018) جمع‌بندی بر روی کلیه متدهای مورد استفاده برای یادگیری هستان‌نگاری انجام گرفته و عنوان شده است که یادگیری هستان‌نگاری یک فرایند گام‌به‌گام است. در این پژوهش، پیکره‌های متنی با استفاده از تکنیک‌های زبانی مانند برچسب‌گذاری اجزای کلام[1]، تجزیه و واژه‌سازی پیش‌پردازش می‌شوند. پس از پیش‌پردازش، اصطلاحات و مفاهیم مربوط به دامنه استخراج می‌شود. این مرحله از تکنیک‌های مختلف پردازش زبان طبیعی مانند تجزیه نحوی، فریم‌های زیرمجموعه‌سازی و استخراج مجموعه کلمات اولیه (سید)[2]، همراه با برخی تکنیک‌ها از حوزه آماری مانند مقدار C/NC، تحلیل متضاد[3]، تحلیل هم‌وقوعی[4]، تحلیل معنایی پنهان[5] و خوشه‌بندی استفاده می‌کند. افزون بر به‌دست آوردن مفهوم‌ها، روابط سلسله‌مراتبی و غیرسلسله‌مراتبی بین این مفاهیم نیز مورد نیاز است. برای این منظور، از ترکیبی از تکنیک‌ها و رویکردهای

1. POS tagging
2. seed
3. contrastive analysis
4. co-occurrence analysis
5. latent semantic analysis (LSA)

آماری و پردازش زبان طبیعی استفاده می‌شود که شامل تحلیل وابستگی، تحلیل واژگانی-نحوی، تحلیل مفهوم‌های رسمی[1]، خوشه‌بندی سلسله‌مراتبی و کاوش قواعد همبسته[2] می‌شود. شایان ذکر است که واژه‌نامهٔ معنایی در هر دو مرحله استخراج اصطلاح/مفهوم و استخراج رابطه مورد استفاده قرار می‌گیرد. در مرحله بعد قوانین با استفاده از برنامه‌نویسی منطق استقرایی[3] شکل می‌گیرند.

همچنین، در پژوهش (Ostani, Azargoon and Cheshmesohrabi 2018) با تبیین متدولوژی‌های موجود و طراحی مدل مفهومی حوزه، یک هستان‌نگاری در حوزه علم‌سنجی توسعه داده شده است.

همان‌طور که بیان شد، در این زیربخش روش‌های مورد استفاده برای ساخت نیمه‌خودکار هستان‌نگاری که در تحقیقات معتبر و به‌روز مورد استفاده قرار می‌گیرند، معرفی شدند. به‌صورت کلی، تلاش کردیم از زیرمجموعه‌ای مناسب از این روش‌ها برای توسعه نیمه‌خودکار هستان‌نگاری بورس و بازارهای مالی فارسی بهره بگیریم که در بخش ۳، عنوان شده است.

۲-۲. کاربرد هستان‌نگاری در بازارهای مالی

در این زمینه، مطابق دانش نویسنده، پژوهش‌های به نسبت محدودی منتشر شده‌اند که در ادامه معرفی می‌شوند.

در پژوهش «وانگ» به‌منظور سنجش تأثیر انواع اخبار بر فعالیت‌های تجاری در بازارهای مالی (از جمله بازار بورس سهام شرکت‌ها) مانند نوسانات قیمت معاملات، حجم معاملات، فراوانی معاملات یک هستان‌نگاری برای آگاهی از اخبار مربوط به ابزارهای مالی ارائه شده است که شامل دو بخش است: بخش اول، یک چارچوب سلسله‌مراتبی برای دانش حوزه ارائه می‌کند که در درجه اول شامل کلاس‌های اخبار، کلاس‌های شرکت‌کنندگان در بازار، کلاس‌های بازار مالی و روابط اولیه بین این کلاس‌هاست. در بخش دوم، از نقشه علّی برای نشان دادن اینکه چگونه طبقات اخبار با حوزه‌هایی از این بازارها مرتبط هستند، استفاده می‌شود.

در این هستان‌نگاری ابتدا درک دانش درباره اخبار در بازارهای مالی طبقه‌بندی می‌شود. برای مثال، یک خبر ممکن است ماهیت‌های جهانی، تجاری، علمی یا سیاسی

1. formal consepts analysis (FCA) 2. associative rules 3. inductive logic programming (ILP)

داشته باشد و یک خبر با هر یک از این ماهیت‌ها ممکن است مربوط به حوزهٔ انرژی، حوزه صنعتی و یا بورس سهام باشد. همین فرایند برای درک دانش بازارهای مالی هم صورت می‌گیرد، برای مثال، این بازارها می‌توانند بازار سهام، بازار فارکس[1] یا بازار کالا باشد. هر کدام از این بازارها می‌توانند حوزه‌های صنعتی متفاوتی داشته باشند و در هر حوزه صنعتی سهام شرکت‌های مختلفی ممکن است وجود داشته باشد.

سرانجام، سعی می‌شود ارتباطات میان این دو سلسله‌مراتب ذکرشده با انواع زیر برقرار گردد:

- ◇ ارتباط رفتاری دارد با؛
- ◇ ممکن است منجر به نوساناتی شود در؛
- ◇ ارزیابی اولیه‌ای دارد از؛
- ◇ پیشنهادی دارد در.

بدین صورت، نقشه ارتباطات علّی بین دو حوزه خبری و مالی برقرار می‌شود.

لازم به ذکر است که در این پژوهش فرایند ساخت هستان‌نگاری (برای اثبات مفهوم) به‌صورت کاملاً دستی و محدود با استفاده از اخبار مربوط به حملات تروریستی و تأثیر آن در بازارهای سرمایه انجام شده است. در انتها نتیجه گرفته شده است که: اول، درک دانش اخبار در حوزه‌های مالی فرایند بسیار پیچیده و جامعی است؛ دوم، ساخت این پایگاه دانش به ساخت مدل‌های معاملاتی بر اساس اخبار در بازارهای ابزار مالی کمک می‌کند؛ سوم، سیستم‌های توسعه داده‌شده (به‌عنوان مثال، سیستم‌های پیش‌بینی قیمت سهام بر اساس اخبار، سیستم‌های حمایتی از فعالان بازار مالی برای جست‌وجوی اخبار مرتبط) در این حوزه توسط هستان‌نگاری تسهیل و پشتیبانی می‌شوند (Wang et al. 2008).

پژوهشی نیز توسط «ژانگ، ژائو و فنگ» به‌منظور کسب دانش صریح از اخبار صورت گرفته است. این مقاله بر اساس مکانیسم کلی دانش حوزه، استدلال دانش و کشف دانش، چارچوبی برای کشف دانش ضمنی از اخبار و به‌کارگیری دانش در پیش‌بینی سهام ایجاد می‌کند. در ابتدا، با توجه به ویژگی‌های حوزه مالی و مکعب مفهومی، ساختار مفهومی صنعت-شرکت- محصول ساخته شده و چارچوب هستان‌نگاری در حوزه‌ای محدود (صنعت پلی‌استر) مطرح می‌شود. سپس، با ساخت هستان‌نگاری

1. foreign exchange market (FOREX)

حوزه مالی، چارچوب مدیریت دانش اخبار مالی پیشنهاد شده است. افزون بر این، با استفاده از ویژگی‌های هستان‌نگاری و قوانین حوزه استخراج‌شده از متن خبر، مکانیسم استدلال دانش اخبار مالی برای دستیابی به کشف دانش اخبار مالی ساخته می‌شود. بدین منظور اخبار را به‌صورت گزاره‌های مقایسه میزان فروش در بازه‌های زمانی، تأثیر تغییر قیمت و وجود مواد اولیه در محصول در نظر گرفته‌اند. در انتها، قوانین استخراج‌شده با استفاده از دانش متخصصان فیلتر شده و به قوانین محدودتری پالایش شده‌اند. سرانجام، با استفاده از قواعد استخراج‌شده توسط متخصصان و هستان‌نگاری پیشنهاد شده، یک مدل مبتنی بر قانون برای پیش‌بینی قیمت سهام ارائه شده که ادعا شده است در تحلیل تجربی صنعت پلی‌استر به‌خوبی عمل می‌کند (Zhang, Zhao and Feng 2019).

در پژوهش «زاک، تئودولیدیس و دیاز» عنوان شده که در زمینه اوراق بهادار مالی و معاملات، کمبود کار برای توسعه هستان‌نگاری به‌ویژه برای بازارهای سهام و کشف موارد کلاهبرداری وجود دارد. این مقاله سعی دارد با ارائه یک چارچوب سیستماتیک برای توسعه و مدیریت یک هستان‌نگاری برای بازارهای سهام، به این شکاف بپردازد. برای دستیابی به این هدف، پژوهشگران استفاده از تکنیک‌هایی را برای گردآوری و ادغام دانش دامنه از منابع موجود در کنار دانش جدیدِ به‌دست‌آمده از تجزیه و تحلیل منابع بدون ساختار بر اساس موارد کلاهبرداری در بازار سهامِ گزارش‌شده توسط کمیسیون بورس اوراق بهادار ارائه می‌کند. چارچوب پیشنهادی و دانش ساخته‌شده از طریق شش مطالعه موردی ک کار را در انواع مختلف برنامه‌های نگهداری و نظارت «بازار سهام» ارزیابی می‌کند (Zaki, Theodoulidis and Diaz 2019).

در پژوهش «داروسکی، یوانوسکی و فرانکه» یک مدل داده برای یک بازار سهام توسعه داده شده است. این مدل مبتنی بر یک هستان‌نگاری ساخته‌شده برای بازار سهام است که موجودیت‌ها و روابط در بازار سهام و سهامداران آن را پوشش می‌دهد. در این مقاله که خیلی مختصر نوشته شده، آمده است که هدف اصلی استفاده از این هستان‌نگاری، اتصال داده‌های خاص بازار سهام و معاملات سهام با اطلاعات عمومی شرکت‌ها از DBPedia است و انتظار می‌رود که ادغام این دو مجموعه داده منجر به آمارهای دقیق‌تر یا شاخص‌های جدیدی شود که به معامله‌گران کمک می‌کند تا تصویر بزرگ بهتری از شرکت‌ها و سهام مورد معامله را مشاهده کنند. برای مثال، با استفاده از اطلاعات مربوط به شهر یا کمپانی در DBPedia می‌توان به نتایجی مانند تعداد بازرگانان در هر شهر، و میانگین عملکرد پرتفوی معاملاتی در هر شهر رسید. به‌منظور اثبات مدل پیشنهادی، با چندین

پرسش‌وجوی SPARQL بر روی مجموعه داده‌هایی که به‌تازگی منتشر شده، گزارشاتی آماده شده و چند ابزار با محوریت کاربر نهایی به‌صورت مفهومی، مفید بودن این هستان‌نگاری را بررسی می‌کنند (Zdraveski, Jovanoski and Franke 2017).

در پژوهش SHARMA et al. (2023) عنوان شده که اخبار مربوط به بازار سهام در دهه اخیر نقشی حیاتی برای کارگزاران یا کاربران دارد و بنابراین، تمرکز این مقاله بر پیش‌بینی احساسات اخبار بازار سهام بر اساس قطبیت و اطلاعات متنی آن‌ها با استفاده از یک شبکه عصبی کانولوشن (CNN) مبتنی بر دانش هستان‌نگاری است. از جزئیات این هستان‌نگاری اطلاعاتی در مقاله بیان نشده است، اما ظاهراً هستان‌نگاری بر اساس اطلاعات کاربران ساخته شده است که بر قطبیت نظرات اثرگذار است.

مقاله «لامپارتر و شنایزلر» روال طراحی یک بازار مبتنی بر هستان‌نگاری را برای تجارت خدمات وب تشریح کرده است. بر اساس تجزیه و تحلیل نیازمندی‌ها برای ایجاد یک بازار خدمات وب، بازاری طراحی شده که این امتیازات را دارد: بازار از یک زبان ارتباطی مبتنی بر هستان‌نگاری استفاده می‌کند که قادر است درخواست، پیشنهادات و توافقات توصیف‌شده معنایی را نشان دهد. افزون بر این، از اطلاعات معنایی برای تقسیم کل بازار به چندین بازار فرعی مستقل استفاده شده است. نشان داده شده که این مفهوم نسبت به مکانیزم موجود کارآمدتر است، یعنی استفاده از دانش پس‌زمینه پیچیدگی کلی را کاهش داده است (Lamparter and Schnizler 2006).

در پژوهش «میشرا» و همکاران آمده است که بازار سهام به عوامل مختلفی مانند قیمت، حجم و غیره بستگی دارد که این عوامل به‌راحتی توسط شاخص‌ها قابل شناسایی هستند و همچنین در ادبیات، شاخص‌های مختلفی در دسترس است که می‌توان از آن‌ها برای کاهش ریسک در بازار سهام هنگام سرمایه‌گذاری یا معامله استفاده کرد. با این حال، هیچ‌یک از شاخص‌ها نمی‌توانند به تنهایی حفاظت کامل در ریسک‌پذیری سرمایه را ارائه دهند. بنابراین، لازم است یا یک شاخص جدید ایجاد شود یا چارچوبی پیشنهاد شود که متشکل از مجموعه‌ای از چند شاخص باشد که بتواند ریسک سرمایه‌گذار را در مقایسه با هر یک از شاخص‌های موجود کاهش دهد. هدف این مقاله پیشنهاد یک چارچوب مبتنی بر هستان‌نگاری است که ویژگی‌های سه شاخص، یعنی باند بولینگر[1]، فیبوناچی[2]، و هایکین-آشی[3] را همزمان ارائه دهد. در این مقاله این شاخص‌ها با یک هستان‌نگاری در قالب کلاس‌ها، ویژگی‌ها، نمونه‌ها ترسیم می‌شوند. نتایج به‌دست‌آمده نشان می‌دهد که چارچوب پیشنهادی به نتایج بهتری نسبت به شاخص‌های موجود دست می‌یابد (Mishra et al. 2023).

1. Bollinger Bands 2. Fibonacci 3. Heikin-Ash

در پژوهش‌های بررسی‌شده، هستان‌نگاری با هدف تحلیل ریسک، تحلیل احساسات، کشف کلاهبرداری یا ساخت بازار با دامنه متفاوت اما مرتبط با موضوع مورد نظر در این پژوهش ساخته شده و مورد استفاده قرار گرفته‌اند. افزون بر این، محققان عمدتاً به‌دنبال استفاده از منابع خبری محدود سعی در ایجاد قوانین حاکم بر بازارهای مالی داشته‌اند. همچنین، مفاهیم موجود در بازار سهام به‌صورت جامع در این پژوهش‌ها استخراج نشده‌اند و از همبستگی بین سهم‌ها و مفاهیمی به‌غیر از اخبار مانند تحلیل شبکه‌های اجتماعی و رویدادهای سیاسی و نیز تمرکز سرمایه‌گذاری در بخش‌های مختلف جغرافیایی و زمانی بحثی به میان نیامده است. افزون بر این، در این پژوهش‌ها فقط روی امکان‌پذیری استفاده از منابع دانشی مانند هستان‌نگاری در پیش‌بینی قیمت سهام کار شده و در عمل از دانش زمینه‌ای در حوزه تحلیل‌های فنی و بنیادی استفاده‌ای نشده است.

جدای از این موضوع حوزه گراف دانش که با حوزه هستان‌نگاری همپوشانی نسبی دارد نیز به‌صورت موردی و جزیره‌ای در پژوهش‌هایی مورد استفاده قرار گرفته است. در نهایت، با توجه به اهمیت ایجاد منابع دانش، تائید اولیه تأثیر آن‌ها در پژوهش‌ها و فقدان پایگاه‌های دانش به‌ویژه در بازار سهام داخلی و اهمیت بسیار بالای تشخیص عوامل و ویژگی‌های تأثیرگذار در فرایند پیش‌بینی قیمت یا جهت حرکت سهام، ضرورت انجام این پژوهش و امکان‌پذیری آن قابل تائید به نظر می‌رسد.

۳. فرایند ساخت هستان‌نگاری بورس و بازارهای مالی فارسی

هستان‌نگاری بورس و بازارهای مالی که در این پژوهش معرفی می‌گردد، شامل ۵۶۵ مفهوم، ۴۹۶ رابطه سلسله‌مراتبی و ۱۳۷ رابطه غیرسلسله‌مراتبی و ۹۳۷ نمونه است که در قالب یکی از بروندادهای پایان‌نامه در «دانشگاه تربیت مدرس» توسط نویسندگان این پژوهش توسعه یافته است. در صدد هستیم که در ادامه کار از این هستان‌نگاری در جهت تزریق دانش عرفی موجود به سیستم‌های یادگیری ماشین در جهت افزایش دقت آن‌ها برای پیش‌بینی روند بازار سهام استفاده کنیم. با توجه به موجود نبودن چنین منبع دانشی برای بازار داخلی، سعی شد با استفاده از دقیق‌ترین روش‌های موجود، به‌صورت برداشت آزاد (نه به‌طور دقیق، از یک پژوهش خاص و گاهی استفاده از ترکیب این روش‌ها) و همچنین با استفاده از روش‌های فراابتکاری برای افزایش دقت و خودکارسازی هرچه بیشتر فرایند یک مجموعه دانشی مناسب در این راستا ایجاد نماییم. مجموعه داده‌های اولیه و گام‌های مورد نیاز برای ساخت این هستان‌نگاری در زیر بخش‌های زیر معرفی می‌شوند.

۳-۱. منبع دادگان

در مرحله نخست، برای ساخت هستان‌نگاری با روش نیمه‌خودکار نیاز به جمع‌آوری متون با حجم و تنوع کافی در حوزه مورد نظر است. با این پیش‌زمینه، متون از منابع مختلف تهیه شد. متون استفاده‌شده برای ساخت هستان‌نگاری عبارت‌اند از:

- ◇ داده‌های نیمه‌ساخت‌یافته ویکی‌پدیا و سایت TSETMC در زمینه بورس؛
- ◇ داده‌های بدون ساختار (متن خام) رسمی مرتبط با بورس؛
- ◇ سایت آموزش بورس کارگزاری مفید؛
- ◇ محتوای آموزشی سایت نظارت، فراچارت و فرانش؛
- ◇ کانال آموزش بورس دکتر کنعانی.

به‌منظور جمع‌آوری این متون با استفاده از کتابخانه‌های مربوط به خزیدن و تست وب به توسعه نرم‌افزاری برای استخراج متون مرتبط با بورس در حوزه آموزش و معرفی بازار بورس اقدام شد. مجموعه این دادگان شامل ۲۴۵۴ جمله و حدود ۶۷۰۰۰ کلمه بوده است که برای ساخت هستان‌نگاری مورد استفاده قرار گرفته شده است.

۳-۲. شرح ابعاد یادگیری

می‌دانیم که قبل از آغاز فرایند یادگیری هستان‌نگاری (به روش نیمه‌خودکار) باید ابعاد یادگیری مشخص شود. ابعاد یادگیری هستان‌نگاری در این پژوهش عبارت‌اند از:

- ◇ عنصر آموختنی: یادگیری ترم‌ها، مفاهیم، روابط و نمونه‌ها با استفاده از منابع یادگیری؛
- ◇ نقطه شروع: دانش اولیه در این مرحله از صفر؛
- ◇ منابع ورودی: منابع زبانی (غیرساخت‌یافته مانند متون بورسی و نیمه‌ساخت‌یافته مانند صفحات ویکی‌پدیا)؛
- ◇ پیش‌پردازش: استفاده از پیش‌پردازش‌های مختلف متنی از قبیل نرمال‌سازی، توکن‌بندی[1]، تجزیه نحوی[2] و برچسب‌زنی‌های اجزای کلام[3] و نقش معنایی[4] و ... با توجه به میزان توسعه هستان‌نگاری؛
- ◇ نتیجه نهایی: یک هستان‌نگاری تخصصی حوزه بورس (خروجی ابزار protégé).

1. tokenization 2. parsing 3. Pos tagging
4. semantic role labeling (SRL)

۳-۳. آماده سازی نیمه‌خودکار هستان‌نگاری بورس و بازارهای مالی

پس از تهیه مجموعه خام اولیه از منابع ذکرشده، ابتدا نیاز به آماده‌سازی و اعمال ابزارهای مختلف پیش‌پردازش متن وجود دارد. سپس، با استفاده از مجموعه‌ای جامع از روش‌های موجود که برداشتی آزاد از پژوهش‌های مرتبط است و همچنین برخی مکاشفه‌های پیشنهادی ساخت هستان‌نگاری آغاز شده است. در این بخش مراحل ساخت هستان‌نگاری شامل استخراج مفاهیم و روابط و نمونه‌ها به تفکیک شرح داده می‌شود.

۳-۳-۱. پیش پردازش

اقدامات انجام‌شده در مرحله پیش‌پردازش به‌شرح زیر است:

- نرمال‌سازی: در این فعالیت اقدام به پاکسازی و مرتب‌کردن متون ورودی شده است تا متن خام برای پردازش زبان طبیعی فراهم شود. از آنجا که برخی از کاراکترها در صفحه‌کلید فارسی بیش از یک یونی‌کد متناظر دارند (مانند «ی» و «ک»)، کاراکترها در این مرحله یکسان می‌شوند. اصلاح نویسه‌ها و جایگزینی نیم‌فاصله برای کلمات مرکب نیز از اقدامات این فعالیت است؛
- جداسازی جملات: پس از نرمال‌سازی، متن به جملات تقسیم می‌شود. این جملات واحدهای ورودی بخش پردازش ماشین قلمداد می‌شوند. ما از علائم نگارشی مانند «.»، «!» و «؟» به‌عنوان جداکننده استفاده می‌کنیم. در ضمن قوانینی در استفاده از این نمادها در کلمات خاص (مانند کلمات اختصاری)، اعداد و URLها در نظر گرفته شده تا از تقسیم‌بندی نادرست جلوگیری شود؛
- توکن‌بندی: به‌صورت کلی دو نوع توکن‌بندی وجود دارد: توکن‌بندی جملات و توکن‌بندی در یک جمله. برای تبدیل هر جمله به دنباله‌ای از توکن‌ها مانند کلمات، علائم نگارشی و اعداد توکن‌بندی انجام می‌شود. از آنجا که فاصله در زبان فارسی یک جداکننده قطعی نیست و ممکن است در یک کلمه رخ دهد یا دو کلمه مجزا بدون هیچ فاصله‌ای در بینشان ظاهر شود، نشانه‌گذاری در این زبان در مقایسه با انگلیسی چالش‌برانگیزتر است. در این مرحله از جعبه‌ابزار هضم برای توکن‌بندی درون جملات استفاده شده است.

۳-۳-۲. استخراج مفاهیم

در این راستا از ترکیبی از روش‌های آماری و زبانی استفاده شده است. منبع مورد

استفاده در این بخش منابع متنی نیمه‌ساخت‌یافته و غیرساخت‌یافته بوده است که در بخش ۳-۱ آمده‌اند.

◇ **عملیات انجام‌شده بر روی منبع غیرساخت‌یافته**

برای استخراج مفاهیم از دو رویکرد زبانی و آماری استفاده شد. به‌صورت جزئی‌تر برای بخش زبانی از معیار TFIDF و برای بخش آماری از شمارش استفاده شد.

معیار TFIDF (تعداد تکرار عبارت در یک سند بخش بر معکوس تعداد اسنادی که عبارت در آن آمده) که برای شناسایی میزان اهمیت یک عبارت در متن مورد استفاده قرار می‌گیرد، به‌صورت ساده به این معناست که یک کلمه یا یک عبارت وقتی مهم است که زیاد تکرار شده باشد، اما در همه اسناد زیاد نباشد. البته، این مفهوم با مفهوم ایست‌واژه‌ها، متفاوت است. ایست‌واژه‌ها لیستی از کلمات پرتکرار هستند که مفهومی ندارند؛ برای مثال، است، با، به، در ... که این لیست از قبل مشخص است و پیش از پردازش از متون حذف می‌گردد. در مورد TFIDF امتیاز کلماتی مانند «ایران» که مثلاً در متن زیاد آمده، اما مفهوم تخصصی حوزه موضوعی را ندارد، نسبت به کلمه «بورس» یا «سهم» کمتر می‌شود. این امتیاز به‌صورت خودکار با استفاده از حاصل‌ضرب دو معیار term frequency (tf) و inverse document frequency (idf) انجام می‌شود که در متون مورد نظر با اعمال رابطه‌های زیر محاسبه شدند (Rajaraman 2011):

$$\mathrm{tf}(t,d) = 0.5 + \frac{0.5 \times \mathrm{f}(t,d)}{\max\{\mathrm{f}(w,d) : w \in d\}}$$

F (t,d) تعداد تکرار کلمه t در سند d (سند هدف)، و maxf (w,d) تعداد پرتکرارترین کلمه در سند d است و

$$\mathrm{idf}(t,D) = \log \frac{N}{|\{d \in D : t \in d\}|}$$

که در آن N تعداد کل اسناد موجود در پیکره است و D بیانگر تعداد اسنادی است که کلمه t در آن‌ها وجود دارد.

در این بخش با توجه به روابط عنوان‌شده روی متون ورودی امتیاز TFIDF کلمات و عبارات (تا ۵ کلمه) محاسبه شد و لیستی از کلمات و عبارات پرتکرار به‌دست آمد. در بخش آماری ابتدا پرتکرارترین کلمات و عبارات (۲-گرام، ۳-گرام و ۴-گرام) متن

استخراج شد. پس از بررسی عبارات استخراج‌شده به نظر می‌رسید که غالب عبارات ۴ و ۵ کلمه‌ای ترکیبی از افعال و یا حروف اضافه با عبارات هستند و برای انتخاب به‌عنوان مفهوم مناسب نیستند. (برای مثال، عبارات «مشمول گره معاملاتی هستند» یا «جلسه معاملاتی متوالی برگزار شد» چندین بار تکرار شده بود). این بود که از لیست کنار گذاشته شدند. از سوی دیگر، با توجه به اینکه تعداد تکرار عبارات به مراتب از کلمات کمتر است (بیشینه تکرار تک‌کلمه‌ها در این پیکره ۲۸۸ بار، اما برای سه‌کلمه‌ای‌ها ۳۰ بار بود)، برای انتخاب بهتر مفاهیم چند ویژگی برای هر مفهوم کاندید مدنظر قرار گرفت:

- ◇ تعداد تکرار؛
- ◇ تعداد کلمات عبارت کاندید؛
- ◇ مکان رخداد جمله دارای عبارت (که به‌دلیل بدون ساختار بودن بخش بزرگی از متن ورودی به نظر بی‌فایده رسید).

پس از این مرحله، تعداد تکرار هر عبارت کاندید با توجه به تعداد کلمات آن با استفاده از تابع نمایی وزن‌دهی شد و عبارات کاندید استخراج شدند.

در این بخش لیست به‌دست‌آمده از هر دو روش TFIDF و بر اساس فرکانس تکرار تجمیع شد و اصلاحاتی روی آن صورت گرفت. برای مثال، با توجه به اینکه همه چندگرام‌های پرتکرار لزوماً خاص‌دامنه نیستند، نیاز بود تا فیلترهایی نیز بر روی آن‌ها اعمال گردد. به‌عنوان مثال، مفاهیم نباید شامل ایست‌واژه، علائم یا رقم باشند (عباراتی چون «حجم مبنا۳ بود، یا «بازار گاوی» می‌گویند» در مجموعه وجود داشت). همچنین، یکسری عبارات پرتکرار که از متون عام استخراج شده بودند، مقایسه و از لیست کاندید حذف شدند.

از سوی دیگر، با توجه به برچسب اجزای کلام عبارات و کلمات کاندید، آن‌هایی که فعل بودند یا قید نیز (مانند باشد و زمانی که فعل و قید هستند) حذف شدند. عبارات پرتکرار عام در صورت رخداد در لیست عبارات کاندید (مانند افزایش، روز، قابل، ایران) حذف شدند.

از این بخش پس از دسته‌بندی مفاهیم و فیلتر مفاهیم، مواردی مانند ۱) مفرد و جمع بودن مانند سهم و سهام، ۲) اضافه داشتن یک کاراکتر یا یک کلمه (مانند نقد و نقدی)، و ۳) هم‌معنا بودن مفاهیم مانند بازار بورس و بازار بورس اوراق بهادار که منظور یک مفهوم است، در نهایت، ۳۰۰ مفهوم انتخاب شد.

نمونه‌ای از مفاهیم استخراج‌شده به‌شرح زیر است:

- سهام
- شرکت
- بازار
- بورس
- اوراق
- محصول
- سرمایه
- کدال

- اوراق بهادار
- افزایش سرمایه
- سهام جایزه
- فروش تبعی
- اختیار فروش
- حق تقدم
- مجمع عمومی
- سازمان بورس

عملیات انجام‌شده بر روی منبع نیمه‌ساخت‌یافته

مهم‌ترین منبع نیمه‌ساخت‌یافته در دسترس صفحات ویکی‌پدیا هستند. با جست‌وجو در این صفحات با تمرکز بر یافتن مفاهیم و نمونه‌های مرتبط با حوزه بازارهای مالی، چند صفحه مشخص از سایت ویکی‌پدیا با عناوین مختلف مرتبط با بورس مانند شخصیت‌ها، شرکت‌ها و ارگان‌های مالی کشور انتخاب شد که دارای دسته‌بندی مناسبی از مفاهیم و نمونه‌های حوزه بورس بود. در جدول ۱، به نمونه‌هایی از آن‌ها اشاره شده است. در ادامه، برای استخراج خودکار مفاهیم از این صفحات منتخب، با توجه به این موضوع که مفاهیم در آن‌ها با لینک و زیرخط مشخص شده بودند، فایل html صفحه ذخیره شد و با استفاده از کدنویسی مفاهیم و سرتیتر آن (کلاس بالاتر) استخراج شد. این فعالیت در واقع، با کمک کد توسعه داده‌شده برای خزیدن وب که در بخش ۳-۱ بدان اشاره شد، صورت گرفت. این مفاهیم با توجه به نیمه‌ساخت‌یافته بودن منبع با دقت بالایی استخراج شدند. (دقت مفاهیم استخراج‌شده بالای ۹۰ درصد است و نقص آن مربوط به بخش‌های کم‌ارتباط‌تر به حوزه مالی بود)

جدول ۱. عناوین انتخاب‌شده برای استخراج داده از ویکی‌پدیا

شخصیت‌ها	ارگان‌های مالی	شرکت‌ها
علی صالح‌آبادی	سازمان بورس و اوراق بهادار	شرکت کارگزاری
علی رحمانی	سازمان کارگزاران بورس اوراق بهادار تهران	شرکت تأمین سرمایه
مصطفی امیدقائمی	بورس اوراق بهادار تهران	شرکت مشاور سرمایه‌گذاری
حسن قالیباف اصل	شرکت فرابورس ایران	صندوق سرمایه‌گذاری مشترک
امیرحمزه مالمیر	وزارتخانه امور اقتصادی و دارایی	شرکت سرمایه‌گذاری
محمدعلی دهقان دهنوی	بورس کالای ایران	شرکت هلدینگ

۳-۳-۳. استخراج روابط

در بخش دوم یادگیری، فرایندهای لازم برای استخراج روابط سلسله‌مراتبی و غیرسلسله‌مراتبی از متن پیاده‌سازی شده است.

◇ **استخراج روابط سلسله‌مراتبی**

برای استخراج روابط سلسله‌مراتبی از روش مبتنی بر الگوها استفاده شده است. ابتدا تعدادی الگو مشابه الگوهای Hearst (1992) مطابق آنچه در زیر می‌آید، انتخاب شده است. سپس، الگوها روی متن اعمال گردیده و ساختارهای موجودی که در قالب آن‌ها گنجانده می‌شدند، به‌دست آمد. گروه‌های اسمی استخراج‌شده برای مرحله انتخاب از طریق یک رابط کاربری به‌عنوان پیشنهاد ارائه شده و عامل انسانی انتخاب نهایی را برای افزوده‌شدن روابط سلسله‌مراتبی به هستان‌نگاری صورت می‌دهد.

الگوهای مورد استفاده شامل مواردی از قبیل زیر هستند:

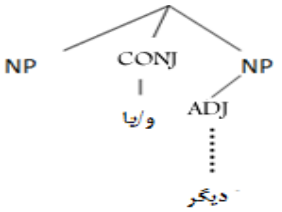
جدول ۲. نمونه الگوهای مورد استفاده برای استخراج روابط سلسله‌مراتبی

NPx شامل NPy	NPx و سایر NPyها
NPx شبیه NPy	NPy از قبیل NPx
این چنین NPxهایی (NPy قبلاً در متن آمده)	NPx نامیده می‌شود NPy

که در آن‌ها NPy به‌عنوان hypernym (کلاس بالاتر - پدر) برای NPx در نظر گرفته می‌شود و رابطه is-a مطابق زیر میان آن‌ها برقرار خواهد بود: Npx Is-A NPy

برای استخراج این الگوها از متن، از درخت تجزیه نحوی جملات استفاده شده است

و در این پروژه از ابزار آماده‌شده بر اساس تجزیه‌گر برکلی[1] استفاده شده است. دقت این ابزار بالای ۸۵ درصد است. شکل درخت زیر مثالی از جایگاه محلی این گروه‌های اسمی برای الگوها را که برای استخراج آن‌ها مورد استفاده قرار گرفته، نشان می‌دهند:



شکل ۲. الگوی نمونه برای استخراج روابط is-a

نمونه‌ای از این موارد در داده‌های بورسی برای الگوی x مانند y، z، f و g: (ارز، طلا، مسکن و خودرو نوعی بازار) عبارت‌اند از:

بازارهای	N-PL-COM
دیگری	PRO
مانند	ADV-EXM
ارز	N-SING-COM
،	DELM
طلا	N-SING-COM
،	DELM
مسکن	N-SING-COM
و	CON
خودرو	N-SING-COM

پیشنهادهای ارائه‌شده به کاربر با توجه به روابط استخراج‌شده میان گروه‌های اسمی از روی الگوها و با لحاظ این نکته صورت گرفته است که برای گروه‌های اسمی که متشکل از بیش از یک واژه باشند، پیشنهادات، افزون بر خود آن گروه‌ها با ترکیب تک‌واژه‌های تشکیل‌دهنده آن‌ها و یا دو واژه، به‌عنوان پدر و فرزند صورت می‌گیرد. همچنین، گروه‌های دو-واژه‌ای بر اساس الگوی صفت-موصوف زیر به‌عنوان روابط

1. Berkeley parser

is-a پیشنهادی ارائه می‌شوند:

- ◇ modifier + noun *is-a* noun؛
- ◇ شرکت نفت *is-a* شرکت.

روش دیگر در این زمینه خوشه‌بندی مفاهیم انتخاب‌شده و انتخاب پدر برای آن‌هاست. به‌عنوان مثال، «پتروشیمی گاز» و «پتروشیمی نفت» هم‌گروه هستند و پدر آن‌ها «پتروشیمی» است. همچنین، با بررسی خروجی تجزیه‌گر وابستگی مشخص شد که انواعی از روابط وجود دارند که روابط سلسله‌مراتبی ایجاد می‌کنند. این روابط مانند amod و POSS. برای مثال، POSS (بازار، فعالان) و POSS (سرمایه، بازار) یعنی بازار، سرمایه و فعالانی دارد و یا amod (مفید، کارگزاری) یعنی مفید نوعی کارگزاری است. این‌گونه روابط نیز به مجموعه اضافه شد.

- ◇ **استخراج روابط غیرسلسله‌مراتبی**

افزون بر روابط is-a برای استخراج سایر روابط، تعداد ۷ فعل پرتکرار (تولیدکردن، فراهم‌کردن، تشخیص دادن، تعلق داشتن، داشتن، ایجاد کردن، و رابطه داشتن) و چند کلیدواژه مانند «جزئی از» و «بخشی از» انتخاب گردید و با استفاده از درخت تجزیه وابستگی و برچسب‌زن نقش معنایی، نشان‌وندهای فاعل و مفعول برای این روابط استخراج شد. بدین ترتیب، رابطه متناظر با آن فعل میان فاعل و مفعول آن ایجاد می‌گردد.

این نشان‌وندها به‌صورت زوج، مفاهیمی که دامنه و برد این روابط را تشکیل می‌دهند، به‌عنوان پیشنهاد به عامل انسانی داده شده است. این پیشنهادات در فرایندی مشابه آنچه در روابط is-a داشتیم، تولید شده‌اند و سپس، به‌صورت نیمه‌خودکار، روابط مناسب تائید شده و به هستان‌نگاری اضافه شده‌اند.

در این مرحله از تجزیه‌گر وابستگی آموزش‌دیده با پیکره اوپسالا[1] استفاده شده است. این تجزیه‌گر با استفاده از ابزار malt optimizer بهینه‌سازی شده است و دقتی در حدود ۸۰ درصد دارد. همچنین، برچسب‌زن نقش معنایی (روی پیکره معنایی نور) آموزش داده شده و دقتی در حدود ۸۰ درصد دارد. این ابزارها بر روی جملاتِ دارای افعال پرتکرار اعمال شدند و نشان‌وندهای مناسب استخراج شده و در مجموعه قرار گرفتند. مانند (گشایش گره بانکی => مرتبط با واقعی شدن نرخ سود).

1. Uppsala Universal Dependencies

برای استخراج روابط غیرسلسله‌مراتبی، افزون بر روش بالا که با انتخاب رابطه از پیش تعیین‌شده و یافتن وابسته‌های آن صورت می‌گرفت، جملاتی از متن ورودی که حداقل دارای دو مفهوم هستند، استخراج شد که با استفاده از تجزیه‌گر وابستگی و برچسب‌زن نقش معنایی روابط بین این مفهوم‌ها استخراج گردید. این کار بر روی مفاهیم استخراج‌شده در بخش ۳-۲ پروژه صورت پذیرفت. برای این کار تلاش شد تا مطابق الگوی مطرح‌شده بررسی شود که آیا این مفاهیم هم‌وقوع به‌عنوان نشان‌وندهای متناظر با فعل واحدی به‌کار رفته‌اند تا رابطه میان آن‌ها تعریف گردد. مجموعه روابط غیرسلسله‌مراتبی استخراج‌شده از متون از قبیل ۳۰ نوع رابطه زیر است.

جدول ۳. نمونه‌ای از انواع روابط غیرسلسله‌مراتبی

مرتبط_با	شامل	درنوردیدن
انجام_می_شود	صادر_می‌کند	رقابت_دارد
انجام_می_دهد	تحت_تأثیر	صورت_گرفتن
ایفای_نقش	است	کارشناسی_کردن
دارند	اصلاح_کردن	گرفتن
بخشی_از	بازگرداندن	متوقف_کردن
نیاز_دارد	ایجاد	خواستن
انجام_می_دهند	بستن	دادن
اهمیت_دارد_برای	بلعیدن	فروختن
منتشر_می_کند	پیش_بینی_کردن	واریز_کردن

سایر تکنیک‌های مورد استفاده

افزون بر موارد فوق، مجموعه‌ای از سایر روش‌ها برای غنی‌سازی هستان‌نگاری بورس مورد نظر قرار گرفت. مثال‌هایی از این تکنیک‌ها که در راستای تکمیل هستان‌نگاری حوزه بورس مورد استفاده قرار گرفت، به‌شرح زیر است:

جدول ٤. نمونه‌ای از تکنیک‌های مکاشفه‌ای مورد استفاده و مثالی از آن‌ها

تکنیک	مثال
استفاده از الگوهای مبتنی بر pos[1] (مثلاً چند اسم عام متوالی)	نماد NN سرمایه NN افزایش NN
استفاده از درخت پارس زایشی[2] برای استخراج درست گروه عبارات در دو سمت یک رابطه سلسله‌مراتبی	[تحلیل تکنیکال و تحلیل بنیادی] IS_a [روش‌های مختلف تحلیل]
استفاده از درخت پارس زایشی برای استخراج گروه عبارات در دو سمت یک فعل خاص در روابط غیررابطه سلسله‌مراتبی	برآورد سود آینده شرکت~ استفاده_می‌کند_از ~قیمت روز سهم
فیلتر مفاهیم پرتکرار با استفاده از بررسی POS آن‌ها	----قیمت تئوریک قیمت
استفاده از NER[3]	PERSON: همتی ORGANIZATION: بانک مرکزی

۳-۳-٤. استخراج نمونه‌ها

برای این بخش از متون نیمه‌ساخت‌یافته استفاده شد. تعدادی صفحه ویکی‌پدیا و نیز سایت TSETMC که دارای متون مناسبی در مورد اشخاص، سازمان‌ها، واحدهای پول و دسته‌بندی حوزه‌های مرتبط بود، انتخاب شد و سپس با استفاده از فایل html آن‌ها و کدنویسی که انجام شد، عبارات لینک‌دار و زیرخط‌دار و پررنگ استخراج گردید. این مرحله نیز با توجه به ورودی مناسب دارای دقت بالایی (۸۵ درصد) بود.

لازم به ذکر است که در این راستا پردازشی نیز بر روی متون غیرساخت‌یافته انجام شد و با استفاده از ابزار NER لیستی از کلمات و عبارات استخراج شد و به نمونه‌ها افزوده گردید.

۳-۳-۵. ویژگی‌ها

در این پژوهش ویژگی‌ها به‌ازای یکسری مفاهیم مانند شخص (نام، اطلاعات شخصی و ...)، شرکت (نام، نوع و ...) و سهم (مانند قیمت، روند) به‌صورت دستی ایجاد و به هستان‌نگاری اضافه گردید.

---

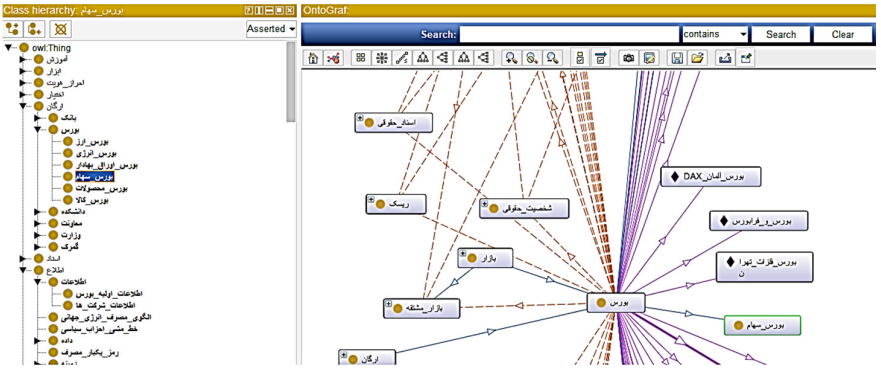
1. part of speech. برچسب‌زن اجزای کلام که وظیفه پیدا کردن نقش نحوی اجزای جمله را بر عهده دارد.

۲. استخراج درخت گرامری اجزای جمله

3. Named Entity Recognition یا شناسایی موجودیت‌های نامدار

۳-۳-۶. تبدیل به فرمت ابزار protégé

همه مفاهیم و روابط استخراج‌شده سرانجام به فرمت owl تبدیل خواهد گردید و امکان مشاهده و اصلاح هستان‌نگاری در محیط protégé نیز بدین ترتیب فراهم خواهد شد. نمایی از این هستان‌نگاری در این ابزار در شکل ۲، آورده شده است.



شکل ۳. نمایی از هستان‌نگاری حاصل در ابزار پروتژ

۳-۳-۷. مشخصات هستان‌نگاری حاصل

هستان‌نگاری حاصل یک هستان‌نگاری خاص دامنه در حوزه بورس است که شامل ۵۶۵ مفهوم، ۴۹۶ رابطه سلسله‌مراتبی و ۱۳۷ رابطه غیر سلسله‌مراتبی و ۹۳۷ نمونه است. لازم به ذکر است که در کلیه مراحل ساخت این هستان‌نگاری از نظر مستقیم کاربر و ارزیابی انسانی بهره گرفته شده است. به همین جهت، دقت مفاهیم و روابط در این هستان‌نگاری بالای ۸۵ درصد است. همچنین، لازم به ذکر است که تعدادی از مفاهیم نیز به‌صورت دستی از مفاهیم مورد نظر متخصصان به مجموعه اضافه شده است.

لازم به ذکر است که با اضافه‌کردن منابع دانشی ساختارمند و بدون ساختار می‌توان غنای هستان‌نگاری مورد نظر را افزایش داد و نیز در حوزه‌های صنعتی خاص مانند حوزه بانکی دانش عمیق‌تری را به سیستم تزریق کرد.

٤. بحث و ارزیابی

مطابق آنچه در Raad & Cruz (2015) آمده، یک هستان‌نگاری را عموماً می‌توان از طریق رویکردهای مختلفی ارزیابی کرد؛ مانند مقایسه با استاندارد مرجع، ارزیابی مبتنی بر داده، ارزیابی مبتنی بر کاربر و ارزیابی کاربردی (وظیفه‌محور). با توجه به دانش

نویسندگان، هستان‌نگاری دیگری برای دامنه بورسی و بازار مالی ایران وجود ندارد. بنابراین، اولین گزینه یعنی مقایسه آن با استاندارد مرجع امکان‌پذیر نیست. افزون بر این، ارزیابی مبتنی بر داده، فرایند مقایسه هستان‌نگاری با داده‌های موجود در مورد دامنه‌ای است که هستان‌نگاری آن را مدل می‌کند. این فرایند دقیقاً رویّه‌ای است که در اینجا برای ساختن هستان‌نگاری دنبال شده است. آخرین مورد نیز به‌صورت کاربردی در آینده پس از استفاده از این هستان‌نگاری در سیستم تحلیل خودکار بورس مورد استفاده و ارزیابی قرار خواهد گرفت که در زمان نگارش این مقاله در دسترس نیست. بنابراین، ارزیابی انسانی یا مبتنی بر کاربر مورد نظر قرار گرفت که در جدول ۴، آورده شده است.

افزون بر این، برای اینکه ارزیابی بهتری از هستان‌نگاری معرفی‌شده ارائه گردد، معیارهای معرفی‌شده در چند پژوهش دیگر را نیز مورد نظر قرار دادیم:

ابتدا، معیار معرفی‌شده توسط Fernández et al. (2009) که بررسی عمق و وسعت هستان‌نگاری است. طبق مطالعات عنوان می‌گردد که در بین معیارهای مختلف عمق و وسعت، مهم‌ترین آن‌ها واریانس وسعت و واریانس عمق هستند و بهترین هستان‌نگاری‌ها به‌طور عام آن‌هایی هستند که مقادیر بالاتری از واریانس‌های عمق و وسعت را در ساختار خود دارند. شایان ذکر است که این معیارها برای مقایسه بیش از یک هستان‌نگاری با هم مناسب هستند و در مورد پژوهش ما دستاوردی ندارند، مگر برای بررسی نسخه‌های آینده که بهبودی ایجاد شده یا خیر. عمق و وسعت هستان‌نگاری فارسی بورس و بازارهای مالی در جدول ۴، آورده شده است.

مورد دوم معیار عنوان‌شده در مرجع Batet & Sanches (2014) است که از فاصله معنایی بین دو کلاس استفاده می‌شود. این معیار مبتنی بر مقایسه مفاهیم و اشتراکات معنایی آن‌هاست. بر این اساس، یک معیار مبتنی بر ویژگی پیشنهاد شده است که فاصله معنایی را بر اساس تعداد پدران در سلسله‌مراتب غیرمشترک بخش بر کل پدران اندازه‌گیری می‌کند. روابط مورد استفاده در شکل ۴، آورده شده‌اند. در این روابط T(c) مجموعه پدرهای یک مفهوم با در نظرگرفتن خودش است.

$$d(c_1,c_2)=\log_2\left(1+\frac{|T(c_1)\cup T(c_2)|-|T(c_1)\cap T(c_2)|}{|T(c_1)\cup T(c_2)|}\right)$$

$$Dispersion(O)=\sqrt{\frac{\sum_{c_i\in C} d(c_i,Root(O))^2}{|C|}}$$

شکل ٤. روابط مورد استفاده برای محاسبه پراکندگی معنایی

همان‌طور که توسط نویسندگان مقاله نتیجه‌گیری شد، مقادیر بالاتر پراکندگی، توزیع مناسب مفاهیم را در هستان‌نگاری نشان می‌دهد. پراکندگی هستی‌نگاری بورس که در جدول ۴، آورده شده، در حدود ۸۲ درصد است که مقدار مناسبی به نظر می‌رسد.

جدول ٤. مقادیر ارزیابی هستان‌نگاری با استفاده از معیارهای عنوان‌شده

معیار ارزیابی	مقدار			
دقت (precision) بر اساس ارزیابی ۲ خبره	۸۵/۷ درصد			
عمق و وسعت هستان‌نگاری بر اساس (Sabou & Motta, 2009)		کمینه	بیشینه	میانگین
	عمق	۲	۹	۴/۲
	وسعت	۲	۷۲	۳/۶
پراکندگی معنایی (Batet & Sanches, 2014)	۸۲ درصد			

با توجه به موارد عنوان‌شده در بخش ارزیابی، هستان‌نگاری ایجادشده با توجه به کیفیت و پوشش مناسب آن می‌تواند در زمان فعلی مجموعه دانشی مناسبی برای استفاده در سیستم‌های مبتنی بر یادگیری ماشین برای استفاده در تحلیل بورس و بازارهای مالی مورد استفاده قرار گیرد. البته، به‌طور قطع، می‌توان با استفاده از منابع ساخت‌یافته و غیرساخت‌یافته بیشتر و دقیق‌تر و همچنین استفاده از نظرات خبرگان این حوزه این منبع را غنی‌تر و دقیق‌تر نمود.

۵. زمینه‌های کاربردی

در این قسمت کاربردهایی که از خروجی این پژوهش ایجاد شده، تشریح می‌شود.

۵-۱. آموزش تحلیلگران بورس و ایجاد پایگاه دانش برای کارگزاری‌ها

قوانین (ارتباطات) مستخرج از هستان‌نگاری و نیز مفاهیم و ویژگی‌های آن‌ها می‌توانند به‌عنوان پایگاه دانشی برای کسب اطلاعات معنادار از دانش نهان موجود در مقالات و منابع متنی موجود در این فضا مورد استفاده قرار گیرد. پیشنهاد قواعد انجمنی معتبر بر اساس آیتم‌های پرتکرار و مبتنی بر دانش برای استفاده و آموزش تحلیلگران حوزه بورس می‌تواند مفید واقع شود. همچنین، در صورت نیاز به ایجاد دایره‌المعارف‌ها در حوزه‌های مالی و سرمایه‌گذاری نیز می‌توان از هستان‌نگاری ایجادشده به شکل گراف دانش و سایر روش‌های بازنمایی استفاده کرد.

۵-۲. بهبود فرایند بازیابی اطلاعات معنایی

در نظام‌های بازیابی اطلاعات استفاده از هستان‌نگاری به‌عنوان راه‌حلی برای مشکلات جست‌وجوی کلیدواژه‌ای و تبدیل جست‌وجوی سنتی به جست‌وجوی معنایی پیشنهاد شده است (Karimi, Babaei, and Hosseini Beheshti 2019). بنابراین، استفاده از آن می‌تواند در بازیابی معنایی مفید باشد. از هستی‌نگاری ایجادشده می‌توان در بسیاری از پایگاه داده‌هایی که دارای فناوری‌های معنایی است و امکان بازیابی معنایی را دارد، استفاده کرد. با توسعه و تحقیق بیشتر بر روی هستی‌نگاری و ایجاد رابط کاربرهای کاربرپسند، امکان پیاده‌سازی هستی‌نگاری در حوزهٔ بازیابی موتورهای کاوش نیز مهیا می‌شود. موتورهای جست‌وجوی مفهومی دارای ویژگی‌های معنایی شامل استفاده از واژگان کنترل‌شده، اجازه پرس‌وجوی ترکیبی با استفاده از مفاهیم، کلیدواژه‌ها، و فراداده‌ها، ارائهٔ نتایج مرتبط با نیاز اطلاعاتی بیان‌شده، رتبه‌بندی و ذخیره بر اساس ارتباط و تعیین محل و نمایش سریع نتایج هستند که ساختار هستان‌نگاری ایجادشده برای این سیستم‌ها امکان دسترسی به اطلاعات حوزه سهام و ارائه این خدمات را میسر میسازد.

۵-۳. کمک به تعیین استراتژی‌های سرمایه‌گذاری افراد در صندوق‌های سرمایه‌گذاری

ارتباطات بین مفاهیم و نمونه‌های استخراج‌شده می‌توانند در کنار تحلیل‌های بنیادی و تحلیل‌های فنی به سرمایه‌گذاران این حوزه برای تصمیم‌گیری کمک شایانی نماید. با استفاده از هستان‌نگاری ایجادشده و امکانات تصویرسازی ارتباطات و مفاهیم می‌توان داشبردهای تحلیلی برای تعیین میزان تأثیر مفاهیم، نمونه‌های مختلف و رویدادها و وقایع مرتبط با این مفاهیم در قیمت و تحلیل روند حرکت قیمت سهام ارائه نمود.

٦. نتیجه‌گیری

در این مقاله، فرایند تولید یک هستان‌نگاری در حوزه بورس و بازارهای مالی به زبان فارسی معرفی شده است. در این پژوهش از ابزارهای مختلف پردازش زبان طبیعی و روش‌های مختلف زبانی و آماری برای ساخت نیمه‌خودکار هستان‌نگاری مورد استفاده قرار گرفته است. همچنین مجموعه‌ای از تکنیک‌های مکاشفه‌ای برای استفاده از خروجی‌های مختلف ابزارها برای غنی‌تر کردن این پایگاه دانش استفاده شده است. در نتیجهٔ این فعالیت‌ها یک هستان‌نگاری خاص‌دامنه در زمینه بورس شامل ۵۶۵ مفهوم، ۴۹۶ رابطه سلسله‌مراتبی و ۱۳۷ رابطه غیرسلسله‌مراتبی و ۹۳۷ نمونه ایجاد شده است.

هدف این است که از این هستان‌نگاری در راستای افزایش دقت و بهره‌وری سیستم‌ها و الگوریتم‌های تحلیل و پیشبینی روند بازار سهام بهره گرفته شود و روال استفاده از دانش عرفی موجود در سیستم‌های یادگیری ماشینی تسهیل گردد. به‌عنوان پژوهش آتی در این زمینه می‌توان به گسترش و افزایش دقت این هستان‌نگاری اشاره نمود.

References

Asim, Muhammad Nabeel, Muhammad Wasim, Muhammad Usman Ghani Khan, Waqar Mahmood, and Hafiza Mahnoor Abbasi. 2018. A Survey of Ontology Learning Techniques and Applications. *Database*: *The Journal of Biological Databases and Curation*. Oxford University Press, Volume 2018, bay101, 1–24.

Batet, Montserrat, and David Sánchez. 2014. “A Semantic Approach for Ontology Evaluation.” In *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*, 138–45. Cyprus: IEEE.

Buitelaar, Paul, Philipp Cimiano, and Bernardo Magnini. 2005. Ontology Learning from Text: An Overview. *Ontology Learning from Text: Methods, Evaluation and Applications.* Amsterdam: IOS Press. 3–12.

Chandrasekaran, B., John R. Josephson, and V. Richard Benjamins. 1999. “What Aro Ontologies, and Why Do We Need Them?” *IEEE Intelligent Systems and Their Applications* 14 (1): 20–26. doi:10.1109/5254.747902.

Chou, Tung Hsiang, John A. Vassar, and Binshan Lin. 2008. Knowledge Management via Ontology Development in Accounting. *Kybernetes* 37 (1): 36–48. doi:10.1108/03684920810850970.

Ding, Ying, and Schubert Foo. 2002. Ontology Research and Development. Part 1 - a Review of Ontology Generation. *Journal of Information Science* 28 (2): 123–36. doi:10.1177/016555150202800204.

Fernández, Miriam, Chwhynny Overbeeke, Marta Sabou, and Enrico Motta. 2009. “What Makes a Good Ontology? A Case-Study in Fine-Grained Knowledge Reuse.” In *The Semantic Web: Fourth Asian Conference, ASWC 2009, Shanghai, China, December 6-9, 2009. Proceedings 4*, 61–75. Springer.

Gruber, Thomas R. 1993. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition* 5 (2): 199–220. doi:10.1006/knac.1993.1008.

Hazman, Maryam, Samhaa R El-Beltagy, and Ahmed Rafea. 2011. “A Survey of Ontology Learning Approaches.” *International Journal of Computer Applications* 22 (9): 36–43.

Hearst, Marti A. 1992. “Automatic Acquisition of Hyponyms from Large Text Corpora.” In *COLING 1992 Volume 2: The 14th International Conference on Computational Linguistics*. Nantes, France.

Heijst, Gertjan van, Sabina Falasconi, Ameen Abu-Hanna, Guus Schreiber, and Mario Stefanelli. 1995. A Case Study in Ontology Library Construction. *Artificial Intelligence In Medicine* 7 (3): 227–55. doi:10.1016/0933-3657(95)00005-Q.

Lamparter, Steffen, and Björn Schnizler. 2006. Trading Services in Ontology-Driven Markets. In *Proceedings of the 2006 ACM Symposium on Applied Computing*, 1679–83. Dijon, France.

Mishra, Ambrish Kumar, Shweta Anand, Narayan C Debnath, and Archana Patel. 2023. “An Ontological Framework for Risk Mitigation in Stock Market.” In *Intelligent Systems and Applications: Select Proceedings of ICISA 2022*, 517–27. Springer.

Ostani, Morteza Mohammadi, Maryam Azargoon, and Mozaffar Cheshmesohrabi. 2018. Methodology of Construction and Design of Ontologies: A Case Study of Scientometrics Field. *Iranian Journal of Information Processing Management* 33 (4): 1793–1822.

Raad, Joe, and Christophe Cruz. 2015. “A Survey on Ontology Evaluation Methods.” In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development, Part of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*. Lisbon, Portugal.

Schreiber, Guus, Bob J Wielinga, and Joost Breuker. 1993. KADS : A Principled Approach to Knowledge-Based System Development. Vol. 11. Academic Press, April 21, 1993. In https://api.semanticscholar.org/CorpusID:57045801.

Shamsfard, Mehrnoush, and Ahmad Abdollahzadeh Barforoush. 2003. “The State of the Art in Ontology Learning: A Framework for Comparison.” *The Knowledge Engineering Review* 18 (4). Cambridge University Press: 293–316. doi:DOI: 10.1017/S0269888903000687.

Sharma, Neha, Mukesh Soni, Sumit Kumar, Rajeev Kumar, Nabamita Deb, and Anurag Shrivastava. 2023. “Supervised Machine Learning Method for Ontology-Based Financial Decisions in the Stock Market.” *ACM Transactions on Asian and Low-Resource Language Information Processing* 22 (5). ACM New York, NY: 1–24.

Uschold, Michael, and Michael Gruninger. 2004. Ontologies and Semantics for Seamless Connectivity. *SIGMOD Record* 33 (4): 58–64. doi:10.1145/1041410.1041420.

Wang, Shanshan, Zhang Zhe, Ye Kang, Huaiqing Wang, and Xiaojian Chen. 2008. An Ontology for Causal Relationships between News and Financial Instruments. *Expert Systems with Applications* 35 (3): 569–80.

Zaki, Mohamed, Babis Theodoulidis, and David Diaz. 2019. “Ontology-Driven Framework for Stock Market Monitoring and Surveillance.” In *HANDBOOK OF GLOBAL FINANCIAL MARKETS: Transformations, Dependence, and Risk Spillovers*, 75–103. Singapore :World Scientific.

Zdraveski, Vladimir, Mihail Jovanoski, and Uwe Franke. 2017. “Stock Market Ontology.” Data-Driven Innovation. 9th International Conference, ICT Innovations 2017, Macedonia 2017, Web proceedings ISSN 1865-0937, 1–7.

Zhang, Lingling, Minghui Zhao, and Zili Feng. 2019. “Research on Knowledge Discovery and Stock Forecasting of Financial News Based on Domain Ontology.” *International Journal of Information Technology and Decision Making* 18 (3): 953–79. doi:10.1142/S0219622019500160.

Zhou, Lina. 2007. Ontology Learning: State of the Art and Open Issues. *Information Technology and Management* 8: 241–52.

محمدحسین ثمنی

متولد ۱۳۶۵، دارای مدرک کارشناسی و کارشناسی ارشد مهندسی فناوری اطلاعات از دانشکده کامپیوتر و فناوری اطلاعات دانشگاه امیرکبیر است. ایشان هم‌اکنون دانشجوی دکتری رشته مهندسی فناوری اطلاعات گرایش مدیریت سیستم‌های اطلاعاتی در دانشکده صنایع دانشگاه تربیت‌مدرس است.

مباحث مربوط به داده‌کاوی، یادگیری ماشین، تحلیل بازارهای مالی و پردازش زبان طبیعی از علایق پژوهشی وی است.

امیر البدوی

متولد ۱۳۴۰، دارای مدرک دکتری در سیستم‌های اطلاعاتی از دانشگاه لندن (LSE) است. ایشان هم‌اکنون استاد مهندسی فناوری اطلاعات در دانشگاه تربیت‌مدرس است.

یادگیری ماشینی و هوش مصنوعی و روش‌شناسی پژوهش در تحقیقات سیستمی از علایق پژوهشی وی است.

پژوهش‌نامه

پردازش و مدیریت اطلاعات