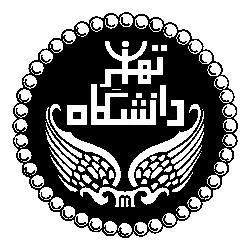
**بسمه‌تعالي**

****

****

**دانشگاهتهران**

**پرديس دانشکده هاي فني**

**دانشکده مهندسي برق و کامپيوتر**

**پيشنهاد و فرم حمايت از پايان‌نامه تحصيلات تكميلي**

**کارشناسي ارشد دکتري**

****

**شماره مرجع**   **:**

\* **شماره مرجع، توسط معاونت پژوهشي پرديس دانشکده‌هاي فنّي هنگام صدور ابلاغ درج خواهد شد.**

**1- خلاصه اطلاعات پايان نامه**

**عنوان پايان نامه به زبان فارسي:**

کشف زیر رخداد در جریان داده‌های خبری

ري

**عنوان پايان نامه به زبان انگليسي:**

Sub event detection in news stream data

**نوع پايان نامه: بنيادي کاربردي توسعه‌اي**

**پرديس/دانشكده: فني دانشكده/گروه: برق و کامپيوتر**

**مقطع تحصيلي : کارشناسي ارشد رشته و گرايش تحصيلي: مهندسی‌ فناوری اطلاعات**

**تاريخ پيشنهاد:7/۷/۹5 تاريخ تصويب:**

**2- اطلاعات اساتيد راهنما و مشاورين**

**نوع مسئوليت**

**نام و نام‌خانوادگي**

**مرتبه علمي**

**محل خدمت**

**امضاء**

**استاد راهنما (مجري)**

**استاد راهنماي دوم (حسب نياز)**

**استاد مشاور**

**استاد مشاور دوم (براي دکتري)**

دکتر آزاده شاکری

استاديار

دانشگاه تهران

**3- اطّلاعات دانشجو**

|  |  |  |
| --- | --- | --- |
| **نام و نام خانوادگی:** مصطفی نورالهی | **شماره دانشجویی:** | 810194237 |
| **رشته و گرايش تحصيلي:** مهندسی‌ فناوری اطلاعات | **دانشکده:** برق و کامپيوتر | **مقطع تحصيلي:** کارشناسي ارشد |
| **پست الكترونيك :**Nourollahi @ut.ac.ir | **تلفن ثابت:** | **تلفن همراه:** 09381417456 |

**4- مشخصات موضوعي پايان نامه**

**تعريف مسأله، هدف و ضرورت اجراي (حداكثر سه صفحه)**

|  |
| --- |
| حجم توليد اطلاعات در سطح جهان روندی تصاعدی و شگفت‌انگیز را طی می‌نماید. استفاده از اطلاعات درصورتی‌که به توليد و ارائه دانش و دانائی منتهی گردد، می‌تواند دستاوردهای مثبتی را به دنبال داشته باشد، در غير این صورت شاهد اتلاف سرمایه‌های انسانی که مهم‌ترین آن عنصر زمان است خواهیم بود. يکی از ویژگی‌های مهم عصر اطلاعات، ميزان توليد، ذخیره‌سازی و نشر اطلاعات در جهان است و یکی از انواع داده‌ها، متون خبری تولیدشده توسط رسانه‌ها است.  حجم عظیم داده‌های خبری تولیدشده توسط رسانه‌های مختلف، اهمیت روش‌هایی برای دسته‌بندی موضوعی، تجمیع خبرهای مشابه ازلحاظ محتوایی و ارائه خبرهای جدید به کاربر به‌صورت آنی، دقیق و پالایش‌شده (بر اساس نیاز کاربر) را آشکار می‌کند.  یکی از پروژه‌های تعریف‌شده در سازمان پروژه‌های تحقیقاتی پیشرفته دفاعی[[1]](#footnote-1) در دهه 90 میلادی، کشف و پیگیری موضوعات مختلف خبری[[2]](#footnote-2) در میان جریانی[[3]](#footnote-3) از مقالات خبری است که هدف اصلی آن، کشف رخدادهای جدید خبری و پیگیری آن‌ها است. این پروژه در سه بخش اصلی تعریف‌شده است:   * قطعه‌بندی[[4]](#footnote-4) جریانی از داده‌های صوتی رسانه‌ها به بخش‌های مجزا و تبدیل صوت به متن * کشف متونی که برای اولین بار در مورد یک رخداد خبری بحث می‌کند * یافتن متون مشابه با داشتن مجموعه‌ای از متون درباره یک رخداد   همان‌طور که اشاره شد، خوشه‌بندی، دسته‌بندی و تحلیل خبرها اجزای اصلی یک سامانه خبری است و این سامانه‌ها، نیازمند الگوریتم‌های بهینه‌ای برای مواجهه به داده‌های پویا به‌منظور خوشه‌بندی، تحلیل و تجمیع زمانی[[5]](#footnote-5) داده‌های خبری است[1]. یکی از جنبه‌های تحلیل خبر، می‌تواند بررسی و تحلیل وقایعی باشد که درون یک رخداد بزرگ‌تر اتفاق می‌افتند همه این رخدادها و زیر رخدادها درباره موضوعی یکسان بحث می‌کنند. رخداد[[6]](#footnote-6) به‌صورت "**چیزی که در زمان و مکان مشخص اتفاق می‌افتد**" تعریف می‌شود. موضوع[[7]](#footnote-7) نیز مفهوم وسیع‌تری از رخداد را شامل می‌شود، یعنی می‌توان "**رخداد**" را یک نمونه از "**موضوع**" در نظر گرفت، به‌عنوان‌مثال زلزله‌ی بم را می‌توان رخدادی برای موضوع حوادث طبیعی برشمرد. به‌طور رسمی، کشف رخداد جدید عبارت است از کشف اولین روایت درباره یک رخداد خاص در میان جریانی از اسناد خبری پیوسته که به لحاظ زمانی مرتب‌شده‌اند[2].  یکی از اهداف مهم سرویس‌های خبری، ارائه جدیدترین اخبار است و فرآیند کشف رخداد جدید می‌تواند به خدمت سرویس‌های خبری درآمده تا با صرف کمترین زمان ممکن، بهترین نتیجه که همانا ارائه اخبار جدید و نادیده گرفتن اخبار تکراری است را به دست آورد؛ اما همان‌طور که در [3] مطرح‌شده است، برخی رخدادهای خبری، دارای ساختار خاص خود می‌باشند، به این صورت که در یک بازه زمانی محدود، تعداد اسناد زیادی درباره یک رخداد خبری منتشر می‌شود و درنتیجه یک رخداد جدید کشف می‌شود. در ادامه، مدتی در مورد رخداد کشف‌شده خبری منتشر نمی‌شود اما بعد از مدت‌زمانی، بازهم اسناد زیادی درباره جنبه‌ای دیگر از آن رخداد خبری منتشر می‌شود. به دلیل مشابهت زیاد اسناد منتشرشده جدید با اسناد قبلی، روش‌های موجود کشف رخداد قادر به تشخیص رخداد جدید نیستند[4] و این در حالی است که اطلاعات جدیدی تولیدشده است که از آن صرف‌نظر می‌شود و این مطلوب نیست.  علاوه بر مسئله مطرح‌شده، نوع دیگری از اطلاعات در جریان یک رخداد خبری زمان‌بندی‌شده[[8]](#footnote-8) مانند یک مسابقه فوتبال و یا یک جشنواره هنری تولید می‌شود. رسانه‌ها تمایل دارند در اسرع وقت جزئیات رخدادهای زمان‌بندی‌شده را به اطلاع کاربر برسانند و علاوه بر این، کاربران نیز در شبکه‌های اجتماعی در مورد اتفاقاتی که طی آن رخداد پیش می‌آید اطلاعاتی را منتشر می‌کنند.  اکثر رخدادها دچار تغییر و تحول می‌شوند و بنابراین، درون هر رخدادی، ممکن است تعداد زیادی زیر رخداد[[9]](#footnote-9) شکل گیرد که رسانه‌ها در مورد این زیر رخدادها، محتوایی تولید می‌کنند که نسبت به گذشته متفاوت بوده و وضعیت جاری رخداد را توصیف می‌کند[4]. زیر رخداد را می‌توان یک بخش مجزا به لحاظ معنایی از یک رخداد کامل دانست[5]. درجایی دیگر [6] زیر رخداد را بخشی از یک رخداد کلی در نظر گرفته است که در یک بازه زمانی کوتاه به‌شدت موردبحث قرارگرفته و سپس به‌صورت تدریجی، این بحث‌ها خاتمه می‌یابد.  در مورد رخدادهایی که بازه زمانی گسترده‌ای را شامل می‌شوند و یا دارای زمان‌بندی شروع و پایان تعیین‌شده‌ای نیستند، تعریف دقیقی را برای زیر رخداد نمی‌توان متصور شد و بایستی با دیدگاه سلسله مراتبی به موضوع نگریست. درعین‌حال می‌توان، دارا بودن اطلاعات جدید درباره یک رخداد کلی را عنصر کلیدی در تشخیص زیر رخداد قلمداد کرد.  راهکار عمومی برای حل این‌ مسئله، نمایش اسناد در قالب مدل فضای برداری[[10]](#footnote-10) و محاسبه فاصله اسناد است[7].  سامانه‌ای که بدین منظور طراحی می‌شود با سه چالش اساسی روبه‌رو است:   * کشف شباهت بین متونی که برای بیان یک رخداد مشابه از کلمات متفاوت استفاده می‌کنند که به دلیل درجه بالای تنوع واژگانی[[11]](#footnote-11) رخ می‌دهد [8]. * افزایش شدید زمان اجرای کشف شباهت بین اسناد به دلیل اضافه شدن اسناد جدید به‌مرورزمان [9] * تولید تعداد زیادی زیر رخداد برای برخی رخدادها که دارای اهمیت یکسان برای کاربر نیست.   برای حل برخی از چالش‌ها در مسئله کشف رخداد راهکارهایی ارائه‌شده است [1]،[7]–[11]، اما ازآنجایی‌که در فرآیند کشف زیر رخداد، شباهت اسناد بسیار بالاست، زیر رخدادها یا به‌عنوان یک رخداد مستقل شناسایی می‌شوند و یا به‌عنوان یک خبر تکراری از آن‌ها صرف‌نظر می‌شود و درنتیجه، این راهکارها در مسئله کشف زیر رخداد موفق عمل نمی‌کنند.  سامانه‌ای که این زیر رخدادها را باوجود تشابه زیاد با اسناد قبلی به‌عنوان یک موجودیت دارای ارزش خبری اعلام کند می‌تواند در کاربردهای مختلف نظیر تولید خط سیر زمانی[[12]](#footnote-12) رخدادهای خبری، بررسی تغییرات قطبیت نظرات کاربران در تحولات یک رویداد خبری و یا به‌عنوان یک دستاورد مستقل در فرآیند کشف رخداد جدید ارزشمند باشد. |
|  |

**روش‌ها و فنون اجرايي طرح**

|  |
| --- |
| به‌طورکلی، در فرآیند کشف رخداد جدید، محاسبه شباهت سند جاری با اسناد قبلی از اهمیت بالایی برخوردار است و می‌توان گفت اکثر روش‌های کشف رخداد، بر این مبنا عمل می‌کنند.  روش‌های موجود در فرآیند کشف زیر رخداد، ماهیت و جزییات رخداد از پیش تعیین‌شده فرض کرده‌اند و علاوه بر این، در اکثر روش‌ها، فرض بر این است که رخداد دارای زمان شروع و پایان از پیش تعیین‌شده است[5]،[13]،[4]،[14]،[6]،[15]،[16]. اما درعین‌حال می‌توان فرض کرد که رخداد از پیش تعیین‌شده نیست که این فرض به واقعیت نزدیک‌تر است، درنتیجه بایستی بتوان یک رخداد جدید را تشخیص داد تا بتوان بعدازآن، زیر رخدادهایی را به آن منصوب کرد، در این مرحله، فنون کشف رخداد قابل‌اعمال است. علاوه بر این، می‌توان از روش‌های مدل‌سازی موضوع[[13]](#footnote-13) نیز به‌منظور تشخیص موضوع اسناد استفاده کرد. در ادامه برای هر رخداد جدید، مجموعه‌ای در نظر گرفته‌شده که به‌مرورزمان، اسنادی به این مجموعه اضافه خواهد شد، سندی در این مجموعه یک زیر رخداد تشخیص داده می‌شود که حاوی اطلاعات جدیدی باشد.  بعدازاینکه برای هر رخداد، مجموعه‌ای از اسناد به دست آمد، با بهره‌گیری از روش‌های استخراج کلمه کلیدی[[14]](#footnote-14)، کلمات نماینده[[15]](#footnote-15) هر رخداد حاصل می‌شود و از این کلمات می‌توان در مراحل بعدی استفاده کرد. با توجه به مسائلی که در بخش قبل در مورد تنوع واژگانی مطرح شد، سعی داریم با بهره‌گیری از روش‌های جدید نمایش طیفی کلمات[[16]](#footnote-16)، به بررسی تأثیر این روش در کاهش اثرات ناشی از تنوع واژگان در اسناد بپردازیم، تأثیر استفاده از این روش در بهبود عملکرد تشخیص رخداد در [12] نشان داده‌شده است. درعین‌حال می‌توان برای هر رخداد خاص، مجموعه‌ی ایست واژه[[17]](#footnote-17) در نظر گرفت تا بتوان فرآیند کشف زیر رخداد را بهبود داد[11]. برای بهبود فرآیند، می‌توان برچسب زمانی سند، موجودیت‌های نامدار[[18]](#footnote-18) استفاده‌شده در سند، عنوان اصلی خبر و ویژگی‌هایی ازاین‌دست را موردمطالعه قرارداد.  یکی از راه‌کارهای ابتکاری[[19]](#footnote-19) به‌منظور کشف زیر رخداد می‌تواند بهره‌گیری از تغییرات نرخ کلمات موضوعی در پیام‌هایی باشد که کاربران در شبکه‌های اجتماعی به اشتراک می‌گذارند، همان‌طور که در [17]،[14]عنوان‌شده است، وقوع یک رخداد منجر به افزایش ناگهانی پیام‌های به اشتراک گذاشته‌شده در شبکه‌های اجتماعی می‌شود. در [4]،[13]،[6]،[5]،[16] از این ویژگی برای تشخیص زیر رخداد استفاده‌شده است، اما نقطه‌ضعف اساسی این روش‌ها این است که مرتباً تغییرات نسبی نرخ کلمات در بازه‌های زمانی مختلف سنجیده شود و این کار، هزینه محاسباتی بالایی دارد، در این پژوهش سعی می‌کنیم تعداد این محاسبات را با حفظ معیارهای کیفی کم کنیم.  به‌طور خلاصه چارچوب این پژوهش و فنون اجرایی آن را می‌توان به‌صورت زیر خلاصه کرد:   * بررسی روش‌های گذشته کشف رخداد جدید و روش‌های کشف زیر رخداد زمان‌بندی‌شده به‌عنوان معیار پایه[[20]](#footnote-20) * مطالعه و بررسی روش‌های استخراج کلمات کلیدی و نیز گسترش کلمات کلیدی رخداد با بهره‌گیری از روش نمایش طیفی * بررسی تأثیر متون تولیدشده توسط کاربران شبکه‌های اجتماعی در فرآیند کشف زیر رخداد * تعیین اهمیت یک زیر رخداد با بهره‌گیری از کمیت و قطبیت نظرات و توییت‌های کاربران و استفاده از روش‌های اندازه‌گیری جنجالی بودن موضوع بدین منظور * بررسی تأثیر ویژگی‌هایی نظیر موجودیت‌های نامدار و برچسب زمانی و نیز تعریف ویژگی‌هایی برای بهبود فرآیند کشف زیر رخداد * تلاش برای کم کردن تعداد تکرارها در فرایند بررسی تغییرات نرخ کلمات کلیدی در شبکه‌های اجتماعی   معیار ارزیابی در این پژوهش، نمودار تعادل خطا[[21]](#footnote-21) یا Curve DET است که در آن، محور افقی، بیانگر نرخ هشدار غلط[[22]](#footnote-22) و محور عمودی بیانگر نرخ از دست رفتن[[23]](#footnote-23) است و عملکرد سیستم در آستانه‌های مختلف برای سنجش شباهت اسناد را توصیف می‌کند. علاوه بر این، از معیارهای دقت[[24]](#footnote-24) و فراخوان[[25]](#footnote-25) نیز در ارزیابی استفاده می‌شود.  مجموعه داده‌ای[[26]](#footnote-26) قابل‌استفاده در این پژوهش مشابه مجموعه داده‌ای استفاده‌شده درزمینه استخراج خط سیر زمانی اخبار است. علاوه بر این می‌توان از مجموعه داده‌ای تویتر نیز استفاده کرد که در [4] استفاده‌شده و در دسترس است. |

**پيشينه تحقيق (همراه با ذکر منابع اساسي)**

|  |
| --- |
| کشف رخداد جدید به‌عنوان یکی از مراحل پروژه تشخیص و پیگیری موضوع، در سال‌های اخیر، موردتوجه پژوهشگران قرارگرفته است، اکثر مسائل در بازیابی اطلاعات بر مبنای پرس و جوی کاربر شکل می‌گیرد و درنتیجه، فرآیند بازیابی و ارائه نتیجه از دانش علایق کاربر بهره می‌گیرد، در سامانه‌های پالایش اطلاعات[[27]](#footnote-27) نیز، با در اختیار داشتن تاریخچه‌ای از درخواست کاربر، اسناد مرتبط به نیاز کاربر در میان انبوهی از اسناد بازیابی می‌شود، اما در کشف رخداد جدید، دانشی ازآنچه که به‌عنوان رخداد جدید رخ خواهد داد وجود ندارد و درنتیجه باید بدون پرس و جوی از پیش تعیین‌شده عمل کرد.  در [18] فرآیند کشف رخداد جدید در چند مرحله انجام می‌شود، ابتدا با فنون استخراج ویژگی و انتخاب ویژگی[[28]](#footnote-28) نمایشی از سند به شکل یک پرس‌وجو تهیه می‌شود و در ادامه، پرس و جوی به‌دست‌آمده با تمام پرس‌وجوهای موجود در حافظه مقایسه می‌شود و بدین ترتیب رخداد جدید کشف می‌شود.  بهره‌گیری از موضوعات عمومی و رده‌بندی اسناد در قالب تعداد محدودی موضوع کلی به‌منظور بهبود فرآیند کشف رخداد موردمطالعه قرارگرفته است. در [11] اسناد در قالب موضوعات گسترده رده‌بندی‌شده و در ادامه، موجودیت‌های نامدار شناسایی و وزن دهی شده و به ازای هر موضوع، مجموعه‌ی ایست واژه ساخته می‌شود و درنهایت، میزان نو بودن هر سند اندازه‌گیری می‌شود.  در [19] عنوان‌شده است که برخی اسناد مرتبط به یک رخداد، تنها در بخش‌هایی از متن دارای اشتراک هستند و ازاین‌رو، هر سند به بخش‌های همپوشان به طول ثابت تقسیم‌شده که این بخش‌بندی ساده در محاسبه شباهت اسناد تأثیر مثبتی داشته است.  علاوه بر روش‌هایی که از مدل فضای برداری استفاده کرده‌اند، تلاش‌هایی برای بهره‌گیری از مدل زبانی[[29]](#footnote-29) در فرآیند کشف رخداد صورت گرفته است اما بهر حال مدل فضای برداری، کاراتر است [2]. علاوه بر این جیمز و همکاران در [2] اذعان داشته‌اند که اسناد نزدیک ازلحاظ زمانی، احتمالاً حاوی اطلاعاتی در مورد رخدادهای مشابه هستند و هرچه فاصله زمانی اسناد بیشتر باشد، احتمال تفاوت در رخداد موردبحث نیز بیشتر است.  حجم عظیم داده‌های خبری که روزانه تولید می‌شود، نیاز به سامانه‌ای که با دقت و سرعت بیشتر رخدادهای جدید را کشف کند را آشکار می‌کند. در [9]،[20]،[10] این مسئله با راهکارهایی نظیر استفاده از ساختار درختی خاص‌منظوره و یا بهره‌گیری از مفهوم پنجره لغزان، بررسی‌شده است.  توییتر به‌عنوان ابزاری برای شناسایی بهنگام رخدادها موردتوجه پژوهشگران قرارگرفته است و اکثر سامانه‌های کشف رخداد از این واقعیت به‌منظور تعیین وقوع زیر رخداد موردعلاقه کاربران درون یک رخداد اصلی استفاده کرده‌اند[4].  بسیاری از پژوهشگران به‌جای کشف رخداد در متون خبری، به کشف رخداد جدید در شبکه‌های اجتماعی و به‌خصوص توییتر علاقه‌مند هستند که به دلیل محبوبیت و ماهیت بهنگام[[30]](#footnote-30) این‌گونه شبکه‌ها، وقوع یک رخداد منجر به اشتراک‌گذاری تعداد زیادی پیام در بازه زمانی کوتاهی می‌شود [14]. به‌طور مثال در [14] تأثیر وقوع زلزله در شبکه اجتماعی موردبررسی قرارگرفته است و مدلی براش پیش‌بینی جهت حرکت طوفان‌ها بر اساس مشاهدات کاربران و توییت‌های[[31]](#footnote-31) آن‌ها ارائه‌شده است، همچنین در [15] از توییتر به‌عنوان یک حسگر[[32]](#footnote-32) در تشخیص رخدادهای طبیعی و اجتماعی استفاده‌شده است.  در [5] یک سامانه کشف زیر رخداد توسعه داده‌شده است، نویسنده از پست‌های تویتر به‌منظور خلاصه‌سازی مسابقات فوتبال آمریکایی استفاده کرده است، روش آن‌ها، ساختار پنهان و واژگان یک بازی فوتبال را با استفاده از مدل پنهان مارکوف[[33]](#footnote-33) یاد می‌گیرد، اما نقطه‌ضعف اصلی روش این است که باید روی رخدادهایی با محدوده زمانی معین اعمال شود.  نیکولاس و همکاران در [21]، روی کشف زیر رخداد با استفاده از حجم توییت‌ها کارکرده‌اند، زمانی که حجم توییت‌ها در مورد یک رخداد خاص، از آستانه‌ای بیشتر شود، یک زیر رخداد تشخیص داده می‌شود. این حجم قبلاً با استفاده از اطلاعاتی در مورد دیگر مسابقات محاسبه‌شده است.  زوبیاگا و همکاران در [16] زیر رخداد را با بهره‌گیری از تفاوت نرخ توییت‌ها در بازه‌های زمانی مشخص به نام قاب زمانی[[34]](#footnote-34) کشف کرده‌اند. زمانی یک زیر رخداد تشخیص داده می‌شود که نرخ توییت در یک قاب زمانی بیشتر از 90 درصد نرخ توییت در بازه‌های قبلی باشد.  در [6] با بهره‌گیری از خوشه‌بندی توییت‌ها بر اساس موجودیت‌هایی که در توییت به آن اشاره‌شده است به کشف زیر رخداد پرداخته‌شده است. نویسنده از یک مدل تلفیقی[[35]](#footnote-35) برای تشخیص زیر رخداد بهره گرفته است.  بهره‌گیری از ویژگی‌های غیر متنی نیز برای کشف لحظات مهم یک رخداد در [13] بررسی‌شده است. در این کار، یک رده‌بند از نوع رگرسیون لجستیک که فقط از نرخ توییت و توییت مجدد استفاده می‌کند، برای تشخیص زیر رخداد استفاده‌شده است.  نمایش گرافی توییت‌های مرتبط به یک رخداد و تقسیم‌بندی آن رخداد به واحدهای زمانی مساوی در [4] به‌خوبی منجر به کشف زیر رخدادها شده است، در این کار، ابتدا با بهره‌گیری از نمایش گرافی مجموعه توییت‌ها در یک بازه زمانی محدود، کلمات پرتکرار و دارای اهمیت به‌دست‌آمده و در ادامه، مقایسه نرخ تغییرات این کلمات مهم به نسبت متوسط نرخ کلمات در بازه‌های قبلی به کشف رخداد جدید منجر می‌شود. |
| **منابع:**  [1] R. Swan and D. Jensen, “TimeMines: Constructing Timelines with Statistical Models of Word Usage,” in *SIGKDD*, 2000.  [2] G. Kumaran and J. Allan, “Text classification and named entities for new event detection,” in *SIGIR*, 2004.  [3] J. B. P. Vuurens and A. P. de Vries, “First Story Detection using Multiple Nearest Neighbors,” in *SIGIR*, 2016.  [4] P. Meladianos, G. Nikolentzos, F. Rousseau, Y. Stavrakas, and M. Vazirgiannis, “Degeneracy-Based Real-Time Sub-Event Detection in Twitter Stream,” in *AAAI*, 2015.  [5] D. Chakrabarti and K. Punera, “Event Summarization using Tweets,” in *ACM international conference on Intelligent User Interfaces*, 2012.  [6] C. Shen, F. Liu, F. Weng, T. Li, and P. Alto, “A Participant-based Approach for Event Summarization Using Twitter Streams,” in *HLT-NAACL*, 2013.  [7] J. Allan, J. Carbonell, G. Doddington, J. P. Yamron, and Y. Yang, “Topic detection and tracking pilot study: Final report,” in *proceedings of the DARPA Broadcast News Transcription and Understanding Workshop*, pp. 194–218.  [8] S. Petrović, M. Osborne, V. Lavrenko, and S. Petrovic, “Using paraphrases for improving first story detection in news and Twitter,” in *NAACL-HLT*, 2012.  [9] K. Zhang, J. Zi, and L. Wu, “New event detection based on indexing-tree and named entity,” in *SIGIR*, 2007.  [10] R.-F. Xu, W.-H. Peng, J. Xu, and X. Long, “On-line new event detection using time window strategy,” in *international Conference on Machine Learning and Cybernetics*, 2011.  [11] Y. Yang, J. Zhang, J. Carbonell, and C. Jin, “Topic-conditioned Novelty Detection,” in *KDD*, 2002, pp. 2–7.  [12] S. Moran, R. Mccreadie, C. Macdonald, and I. Ounis, “Enhancing First Story Detection using Word Embeddings,” in *SIGIR*, 2016.  [13] F. Chierichetti, J. Kleinberg, and R. Kumar, “Event Detection via Communication Pattern Analysis,” *AAAI*. 2014.  [14] T. Sakaki, “Earthquake Shakes Twitter Users : Real-time Event Detection by Social Sensors,” in *www*, 2010, pp. 851–860.  [15] S. Zhao, L. Zhong, J. Wickramasuriya, and V. Vasudevan, “Human as Real-Time Sensors of Social and Physical Events: A Case Study of Twitter and Sports Games,” *CoRR, abs/1106.4300*, 2011.  [16] A. Zubiaga, D. Spina, E. Amigó, and J. Gonzalo, “Towards real-time summarization of scheduled events from twitter streams,” in *Proceedings of the 23rd ACM conference on Hypertext and social media*, 2012.  [17] B. Tsolmon and K. S. Lee, “Extracting social events based on timeline and user reliability analysis on twitter,” in *Proceedings of the 15th International Conference on Computational Linguistics and Intelligent Text Processing*, 2014.  [18] J. Allan, R. Papka, and V. Lavrenko, “On-line New Event Detection,” in *SIGIR*, 1998.  [19] T. Brants and F. Chen, “A System for new event detection,” in *SIGIR*, 2003.  [20] G. Luo, C. Tang, and P. S. Yu, “Resource-adaptive real-time new event detection,” in *SIGMOD*, 2007, p. 497.  [21] J. Nichols, J. Mahmud, C. Drews, H. Rd, and S. Jose, “Summarizing Sporting Events Using Twitter,” in *IUI*, 2012. |

**5- مصوبه شوراي پژوهشي و تحصيلات تکميلي‌ ‌دانشکده مهندسي برق و کامپيوتر**

**5-1- فرم پيشنهاد و حمايت از پايان‌نامه در تاريخ ………………… در شوراي پژوهشي و تحصيلات تکميلي‌**

**‌دانشکده /گروه………………… مطرح و نظر شورا به شرح زير اعلام مي‌شود:**

**تصويب شد نياز به اصلاح دارد به تصويب نرسيد**

**5-2- عنوان طرح جامع تحقيقات استاد راهنما:** سیستم‌های اطلاعاتی و محیط‌های هوشمند

**امضاء استاد راهنما**

**5-3- آيا پايان نامه پيشنهادي مرتبط با طرح جامع تحقيقات استاد راهنما/مشاور/گروه آموزشي/**

**دانشكده مي باشد:**

**بلي خير**

**امضاء رئيس / معاون پژوهشي و تحصيلات تکميلي ‌دانشکده مهندسي برق و کامپيوتر**

**شماره:**

**تاريخ:**

**معاون محترم آموزشي و تحصيلات تکميلي پرديس دانشکده هاي فني**

**با سلام و احترام,**

**فرم پيشنهاد و حمايت از پايان‌نامه کارشناسي ارشد آقاي** مصطفی نورالهی **با عنوان** کشف زیر رخداد در متون خبری **به راهنمايي خانم دکتر آزاده شاکری در شوراي پژوهشي و تحصيلات تکميلي دانشکده مهندسي برق و کامپيوتر مورخ .................... به تصويب رسيد.**

**خواهشمند است دستور فرماييد اقدامات مقتضي انجام شود.**

**امضاء رئيس / معاون پژوهشي و تحصيلات تکميلي دانشکده مهندسي برق و کامپيوتر**

**شماره:**

**تاريخ:**

**معاون محترم پژوهشي پرديس دانشکده هاي فني**

**با سلام و احترام ,**

**به پيوست فرم پيشنهاد و حمايت از پايان‌نامه تحصيلات تكميلي با مشخصات مذکور که به تصويب شوراي پژوهشي وتحصيلات تکميلي دانشکده مهندسي برق و کامپيوتر رسيده است، جهت دستور اقدام مقتضي تقديم مي‌شود.**

**امضاء معاون آموزشي و تحصيلات تکميلي پرديس دانشکده هاي فنّي**

**رونوشت: معاون محترم پژوهشي و تحصيلات تکميلي دانشکده مهندسي برق و کامپيوتر: جهت اطلاع و پيگيري**

1. DARPA [↑](#footnote-ref-1)
2. topic detection and tracking [↑](#footnote-ref-2)
3. stream [↑](#footnote-ref-3)
4. segmentation [↑](#footnote-ref-4)
5. temporal aggregation [↑](#footnote-ref-5)
6. event [↑](#footnote-ref-6)
7. topic [↑](#footnote-ref-7)
8. Scheduled [↑](#footnote-ref-8)
9. Sub event [↑](#footnote-ref-9)
10. Vector space model [↑](#footnote-ref-10)
11. Lexical variation [↑](#footnote-ref-11)
12. Storyline [↑](#footnote-ref-12)
13. Topic modeling [↑](#footnote-ref-13)
14. Keyword extraction [↑](#footnote-ref-14)
15. Representative [↑](#footnote-ref-15)
16. Word embedding [↑](#footnote-ref-16)
17. Stop word [↑](#footnote-ref-17)
18. Named entity [↑](#footnote-ref-18)
19. heuristic [↑](#footnote-ref-19)
20. Baseline [↑](#footnote-ref-20)
21. detection error tradeoff [↑](#footnote-ref-21)
22. False alarm rate [↑](#footnote-ref-22)
23. Missed rate [↑](#footnote-ref-23)
24. precision [↑](#footnote-ref-24)
25. recall [↑](#footnote-ref-25)
26. dataset [↑](#footnote-ref-26)
27. Information Filtering [↑](#footnote-ref-27)
28. Feature selection [↑](#footnote-ref-28)
29. Language model [↑](#footnote-ref-29)
30. Real-time [↑](#footnote-ref-30)
31. tweet [↑](#footnote-ref-31)
32. sensor [↑](#footnote-ref-32)
33. Hidden Markov Model [↑](#footnote-ref-33)
34. Time Frame [↑](#footnote-ref-34)
35. Mixture Model [↑](#footnote-ref-35)