

بسمه تعالی



دانشگاه تهران

پردیس دانشکده های فنی
دانشکده مهندسی برق و کامپیوتر

پیشنهاد و فرم حمایت از پایان نامه تحصیلات تکمیلی

☐

دکتری

☐

کارشناسی ارشد

شماره مرجع * :

* شماره مرجع، توسط معاونت پژوهشی پردیس دانشکده های فنی هنگام صدور ابلاغ درج خواهد شد.

۱- خلاصه اطلاعات پایان نامه

عنوان پایان نامه به زبان فارسی:	
کشف زیر رخداد در جریان داده‌های خبری	
عنوان پایان نامه به زبان انگلیسی:	
Sub event detection in news stream data	
نوع پایان نامه:	<input type="checkbox"/> بنیادی <input checked="" type="checkbox"/> کاربردی <input checked="" type="checkbox"/> توسعه‌ای
پردیس/دانشکده: فنی	دانشکده/گروه: برق و کامپیوتر
مقطع تحصیلی: کارشناسی ارشد	رشته و گرایش تحصیلی: مهندسی فناوری اطلاعات
تاریخ پیشنهاد: ۹۵/۷/۷	تاریخ تصویب:

۲- اطلاعات اساتید راهنما و مشاورین

نوع مسئولیت	نام و نام خانوادگی	مرتبه علمی	محل خدمت	امضاء
استاد راهنما (مجری)	دکتر آزاده شاکری	استادیار	دانشگاه تهران	
استاد راهنمای دوم (حسب نیاز)				
استاد مشاور				
استاد مشاور دوم (برای دکتری)				

۳- اطلاعات دانشجو

نام و نام خانوادگی: مصطفی نوراللهی	شماره دانشجویی: ۸۱۰۱۹۴۲۳۷
رشته و گرایش تحصیلی: مهندسی فناوری اطلاعات	مقطع تحصیلی: کارشناسی ارشد
پست الکترونیک: Nourollahi @ut.ac.ir	تلفن همراه: ۰۹۳۸۱۴۱۷۴۵۶
	تلفن ثابت:

۴- مشخصات موضوعی پایان نامه

تعریف مسأله، هدف و ضرورت اجرای (حداکثر سه صفحه)

حجم تولید اطلاعات در سطح جهان روندی تصاعدی و شگفت‌انگیز را طی می‌نماید. استفاده از اطلاعات در صورتی که به تولید و ارائه دانش و دانائی منتهی گردد، می‌تواند دستاوردهای مثبتی را به دنبال داشته باشد، در غیر این صورت شاهد اتلاف سرمایه‌های انسانی که مهم‌ترین آن عنصر زمان است خواهیم بود. یکی از ویژگی‌های مهم عصر اطلاعات، میزان تولید، 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، تجمع خبرهای مشابه از لحاظ محتوایی و ارائه خبرهای جدید به کاربر به صورت آنی، دقیق و پالایش‌شده (بر اساس نیاز کاربر) را آشکار می‌کند.

یکی از پروژه‌های تعریف‌شده در سازمان پروژه‌های تحقیقاتی پیشرفته دفاعی^۱ در دهه ۹۰ میلادی، کشف و پیگیری موضوعات مختلف خبری^۲ در میان جریانی^۳ 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^۴ 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 تجمع زمانی^۵ داده‌های خبری است [۱]. 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^۶ به صورت "چیزی که در زمان و مکان مشخص اتفاق می‌افتد" تعریف می‌شود. موضوع^۷ نیز مفهوم وسیع‌تری از رخداد را شامل می‌شود، یعنی می‌توان "رخداد" را یک نمونه از "موضوع" در نظر گرفت، به عنوان مثال 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 [۲].

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 [۳] مطرح‌شده است، برخی رخدادها 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 در نتیجه یک رخداد جدید کشف می‌شود. در ادامه، مدتی در مورد رخداد کشف‌شده خبری منتشر نمی‌شود اما بعد از مدت زمانی، باز هم 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 [۴] و این در حالی است که اطلاعات جدیدی تولیدشده است که از آن صرف‌نظر می‌شود و این مطلوب نیست.

¹ DARPA

² topic detection and tracking

³ stream

⁴ segmentation

⁵ temporal aggregation

⁶ event

⁷ topic

علاوه بر مسئله مطرح شده، نوع دیگری از اطلاعات در جریان یک رخداد خبری زمان بندی شده^۸ مانند یک مسابقه فوتبال و یا یک جشنواره هنری تولید می شود. رسانه ها تمایل دارند در اسرع وقت جزئیات رخدادها را زمان بندی شده را به اطلاع کاربر برسانند و علاوه بر این، کاربران نیز در شبکه های اجتماعی در مورد اتفاقاتی که طی آن رخداد پیش می آید اطلاعاتی را منتشر می کنند.

اکثر رخدادها دچار تغییر و تحول می شوند و بنابراین، درون هر رخدادی، ممکن است تعداد زیادی زیر رخداد^۹ شکل گیرد که رسانه ها در مورد این زیر رخدادها، محتوایی تولید می کنند که نسبت به گذشته متفاوت بوده و وضعیت جاری رخداد را توصیف می کند [۴]. زیر رخداد را می توان یک بخش مجزا به لحاظ معنایی از یک رخداد کامل دانست [۵]. درجایی دیگر [۶] زیر رخداد را بخشی از یک رخداد کلی در نظر گرفته است که در یک بازه زمانی کوتاه به شدت مورد بحث قرار گرفته و سپس به صورت تدریجی، این بحث ها خاتمه می یابد.

در مورد رخدادهایی که بازه زمانی گسترده ای را شامل می شوند و یا دارای زمان بندی شروع و پایان تعیین شده ای نیستند، تعریف دقیقی را برای زیر رخداد نمی توان متصور شد و بایستی با دیدگاه سلسله مراتبی به موضوع نگریست. درعین حال می توان، 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^{۱۰} و محاسبه فاصله اسناد است [۷].

سامانه ای که بدین منظور طراحی می شود با سه چالش اساسی روبه رو است:

- کشف شباهت بین متونی که برای بیان یک رخداد مشابه از کلمات متفاوت استفاده می کنند که به دلیل درجه بالای تنوع واژگانی^{۱۱} رخ می دهد [۸].

- افزایش شدید زمان اجرای کشف شباهت بین اسناد به دلیل اضافه شدن اسناد جدید به مرور زمان [۹]

- تولید تعداد زیادی زیر رخداد برای برخی رخدادها که دارای اهمیت یکسان برای کاربر نیست.

برای حل برخی از چالش ها در مسئله کشف رخداد راهکارهایی ارائه شده است [۷]، [۱۱]، [۱۰]، اما از آنجایی که در فرآیند کشف زیر رخداد، شباهت اسناد بسیار بالاست، زیر رخدادها یا به عنوان یک رخداد مستقل شناسایی می شوند و یا به عنوان یک خبر تکراری از آن ها صرف نظر می شود و در نتیجه، این راهکارها در مسئله کشف زیر رخداد موفق عمل نمی کنند.

سامانه ای که این زیر رخدادها را با وجود تشابه زیاد با اسناد قبلی به عنوان یک موجودیت دارای ارزش خبری اعلام کند می تواند در کاربردهای مختلف نظیر تولید خط سیر زمانی^{۱۲} رخدادهای خبری، بررسی تغییرات قطبیت نظرات کاربران در تحولات یک رویداد خبری و یا به عنوان یک دستاورد مستقل در فرآیند کشف رخداد جدید ارزشمند باشد.

⁸ Scheduled

⁹ Sub event

¹⁰ Vector space model

¹¹ Lexical variation

¹² Storyline

به‌طور کلی، 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 [۵]، [۱۳]، [۴]، [۱۴]، [۶]، [۱۵]، [۱۶]. اما در عین حال می‌توان فرض کرد که رخداد از پیش تعیین‌شده نیست که این فرض به واقعیت نزدیک‌تر است، در نتیجه 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^{۱۳} نیز به‌منظور تشخیص موضوع اسناد استفاده کرد. در ادامه برای هر رخداد جدید، مجموعه‌ای در نظر گرفته‌شده که به‌مرور زمان، 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^{۱۴}، کلمات نماینده^{۱۵} 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^{۱۶}، 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 [۱۲] نشان داده‌شده است. در عین حال می‌توان برای هر رخداد خاص، مجموعه‌ای ایست‌واژه^{۱۷} در نظر گرفت تا بتوان فرآیند کشف زیر رخداد را بهبود داد [۱۱]. برای بهبود فرآیند، می‌توان برچسب زمانی سند، موجودیت‌های نامدار^{۱۸} استفاده‌شده در سند، عنوان اصلی خبر و ویژگی‌هایی از این دست را مورد مطالعه قرار داد. یکی از راه کارهای ابتکاری^{۱۹} 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 [۱۷]، [۱۴] عنوان‌شده است، وقوع یک رخداد منجر به افزایش ناگهانی پیام‌های به اشتراک گذاشته‌شده در شبکه‌های اجتماعی می‌شود. در [۴]، [۱۳]، [۶]، [۵]، [۱۶] از این ویژگی برای تشخیص زیر رخداد استفاده‌شده است، اما نقطه ضعف 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^{۲۰}
- 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

¹³ Topic modeling

¹⁴ Keyword extraction

¹⁵ Representative

¹⁶ Word embedding

¹⁷ Stop word

¹⁸ Named entity

¹⁹ heuristic

²⁰ Baseline

- 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^{۲۱} یا DET Curve است که در آن، محور افقی، بیانگر نرخ هشدار غلط^{۲۲} و محور عمودی بیانگر نرخ از دست رفتن^{۲۳} است و عملکرد سیستم در آستانه‌های مختلف برای سنجش شباهت اسناد را توصیف می‌کند. علاوه بر این، از معیارهای دقت^{۲۴} و فراخوان^{۲۵} نیز در ارزیابی استفاده می‌شود. مجموعه داده‌ای^{۲۶} قابل استفاده در این پژوهش مشابه مجموعه داده‌ای استفاده شده در زمینه استخراج خط سیر زمانی اخبار است. علاوه بر این می‌توان از مجموعه داده‌ای تویتر نیز استفاده کرد که در [۴] استفاده شده و در دسترس است.

پیشینه تحقیق (همراه با ذکر منابع اساسی)

کشف رخداد جدید به عنوان یکی از مراحل پروژه تشخیص و پیگیری موضوع، در سال‌های اخیر، مورد توجه پژوهشگران قرار گرفته است، اکثر مسائل در ارزیابی اطلاعات بر مبنای پرس و جوی کاربر شکل می‌گیرد و در نتیجه، فرآیند ارزیابی و ارائه نتیجه از دانش علایق کاربر بهره می‌گیرد، در سامانه‌های پالایش اطلاعات^{۲۷} نیز، با در اختیار داشتن تاریخچه‌ای از درخواست کاربر، اسناد مرتبط به نیاز کاربر در میان انبوهی از اسناد ارزیابی می‌شود، اما در کشف رخداد جدید، دانشی از آنچه که به عنوان رخداد جدید رخ خواهد داد وجود ندارد و در نتیجه باید بدون پرس و جوی از پیش تعیین شده عمل کرد.

در [۱۸] فرآیند کشف رخداد جدید در چند مرحله انجام می‌شود، ابتدا با فنون استخراج ویژگی و انتخاب ویژگی^{۲۸} نمایشی از سند به شکل یک پرس‌وجو تهیه می‌شود و در ادامه، پرس و جوی به دست آمده 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 به منظور بهبود فرآیند کشف رخداد مورد مطالعه قرار گرفته است. در [۱۱] اسناد در قالب موضوعات گسترده رده‌بندی شده و در ادامه، موجودیت‌های نامدار شناسایی و وزن دهی شده و به ازای هر موضوع، مجموعه‌ای ایست واژه ساخته می‌شود و در نهایت، میزان نو بودن هر سند اندازه‌گیری می‌شود.

در [۱۹] عنوان شده است که برخی اسناد مرتبط به یک رخداد، تنها در بخش‌هایی از متن دارای اشتراک هستند و از این رو، هر سند به بخش‌های همپوشان به طول ثابت تقسیم شده 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^{۲۹} در فرآیند کشف رخداد صورت گرفته است اما بهر حال مدل فضای برداری، کارا تر است [۲]. علاوه بر این جیمز و همکاران در

²¹ detection error tradeoff

²² False alarm rate

²³ Missed rate

²⁴ precision

²⁵ recall

²⁶ dataset

²⁷ Information Filtering

²⁸ Feature selection

²⁹ Language model

[۲] اذعان داشته‌اند که اسناد نزدیک از لحاظ زمانی، احتمالاً حاوی اطلاعاتی در مورد رخدادهای مشابه هستند و هرچه فاصله زمانی اسناد بیشتر باشد، احتمال تفاوت در رخداد مورد بحث 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 [۹]، [۲۰]، [۱۰] این مسئله با راهکارهایی نظیر استفاده از ساختار درختی خاص‌منظوره و یا بهره‌گیری از مفهوم پنجره لغزان، بررسی شده است.

توییت‌ها به‌عنوان ابزاری برای شناسایی بهنگام رخدادها مورد توجه پژوهشگران قرار گرفته است و اکثر سامانه‌های کشف رخداد از این واقعیت به‌منظور تعیین وقوع زیر رخداد مورد علاقه کاربران درون یک رخداد اصلی استفاده کرده‌اند [۴]. بسیاری از پژوهشگران به‌جای کشف رخداد در متون خبری، به کشف رخداد جدید در شبکه‌های اجتماعی و به‌خصوص توییت‌ها علاقه‌مند هستند که به دلیل محبوبیت و ماهیت بهنگام^{۳۰} این‌گونه شبکه‌ها، وقوع یک رخداد منجر به اشتراک‌گذاری تعداد زیادی پیام در بازه زمانی کوتاهی می‌شود [۱۴]. به‌طور مثال در [۱۴] تأثیر وقوع زلزله در شبکه اجتماعی مورد بررسی قرار گرفته است و مدلی برآش پیش‌بینی جهت حرکت طوفان‌ها بر اساس مشاهدات کاربران و توییت‌های^{۳۱} آن‌ها ارائه شده است، همچنین در [۱۵] از توییت‌ها به‌عنوان یک حسگر^{۳۲} در تشخیص رخدادهای طبیعی و اجتماعی استفاده شده است.

در [۵] یک سامانه کشف زیر رخداد توسعه داده شده است، نویسندگان از پست‌های توییت‌ها به‌منظور خلاصه‌سازی مسابقات فوتبال آمریکایی استفاده کرده است، روش آن‌ها، ساختار پنهان و واژگان یک بازی فوتبال را با استفاده از مدل پنهان مارکوف^{۳۳} یاد می‌گیرد، اما نقطه ضعف اصلی روش این است که باید روی رخدادهایی با محدوده زمانی معین اعمال شود.

نیکولاس و همکاران در [۲۱]، روی کشف زیر رخداد با استفاده از حجم توییت‌ها کار کرده‌اند، 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 محاسبه شده است.

زوبیگا و همکاران در [۱۶] زیر رخداد را با بهره‌گیری از تفاوت نرخ توییت‌ها در بازه‌های زمانی مشخص به نام قاب زمانی^{۳۴} کشف کرده‌اند. زمانی یک زیر رخداد تشخیص داده می‌شود که نرخ توییت در یک قاب زمانی بیشتر از ۹۰ درصد نرخ توییت در بازه‌های قبلی باشد.

در [۶] با بهره‌گیری از خوشه‌بندی توییت‌ها بر اساس موجودیت‌هایی که در توییت به آن اشاره شده است به کشف زیر رخداد پرداخته شده است. نویسندگان از یک مدل تلفیقی^{۳۵} برای تشخیص زیر رخداد بهره گرفته است.

بهره‌گیری از ویژگی‌های غیر متنی نیز برای کشف لحظات مهم یک رخداد در [۱۳] بررسی شده است. در این کار، یک رده‌بند از نوع رگرسیون لجستیک که فقط از نرخ توییت و توییت مجدد استفاده می‌کند، برای تشخیص زیر رخداد استفاده شده است.

نمایش گرافی توییت‌های مرتبط به یک رخداد و تقسیم‌بندی آن رخداد به واحدهای زمانی مساوی در [۴] 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 به‌دست آمده و در ادامه، مقایسه نرخ تغییرات این کلمات مهم به نسبت متوسط نرخ کلمات در بازه‌های قبلی به کشف رخداد جدید منجر می‌شود.

³⁰ Real-time

³¹ tweet

³² sensor

³³ Hidden Markov Model

³⁴ Time Frame

³⁵ Mixture Model

- [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. Stavarakas, 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. Tsofmon 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.

۵- مصوبه شورای پژوهشی و تحصیلات تکمیلی دانشکده مهندسی برق و کامپیوتر

۵-۱- فرم پیشنهاد و حمایت از پایان نامه در تاریخ در شورای پژوهشی و تحصیلات تکمیلی دانشکده /گروه مطرح و نظر شورا به شرح زیر اعلام می شود:

☐ به تصویب نرسید

☐ نیاز به اصلاح دارد

☐ تصویب شد

۵-۲- عنوان طرح جامع تحقیقات استاد راهنما: سیستم های اطلاعاتی و محیط های هوشمند

امضاء استاد راهنما

۵-۳- آیا پایان نامه پیشنهادی مرتبط با طرح جامع تحقیقات استاد راهنما/مشاور/گروه آموزشی/ دانشکده می باشد:

☐ خیر

☐ بلی

امضاء رئیس / معاون پژوهشی و تحصیلات تکمیلی دانشکده مهندسی برق و کامپیوتر

شماره:

تاریخ:

معاون محترم آموزشی و تحصیلات تکمیلی پردیس دانشکده های فنی
با سلام و احترام،

فرم پیشنهاد و حمایت از پایان نامه کارشناسی ارشد آقای مصطفی نورالهی با عنوان کشف زیر رخداد در متون
خبری به راهنمایی خانم دکتر آزاده شاکری در شورای پژوهشی و تحصیلات تکمیلی دانشکده مهندسی برق و کامپیوتر
مورخ به تصویب رسید.
خواهشمند است دستور فرمایید اقدامات مقتضی انجام شود.

امضاء رئیس / معاون پژوهشی و تحصیلات تکمیلی دانشکده مهندسی برق و کامپیوتر

شماره:

تاریخ:

معاون محترم پژوهشی پردیس دانشکده های فنی
با سلام و احترام ،

به پیوست فرم پیشنهاد و حمایت از پایان نامه تحصیلات تکمیلی با مشخصات مذکور که به تصویب شورای پژوهشی و تحصیلات
تکمیلی دانشکده مهندسی برق و کامپیوتر رسیده است، جهت دستور اقدام مقتضی تقدیم می شود.

امضاء معاون آموزشی و تحصیلات تکمیلی پردیس دانشکده های فنی

رونوشت: معاون محترم پژوهشی و تحصیلات تکمیلی دانشکده مهندسی برق و کامپیوتر: جهت اطلاع و پیگیری