علوم رایانش و فناوری اطلاعات
نشریه علمی انجمن کامپیوتر ایران
مجلد ۱۸، شماره ۲، ۱۳۹۹
صفحات ۴۰- ۴۹
**مقاله عادی**

# ارتقاء فرآیند مطالعه نگاشتی نظام‌مند از طریق یک رویکرد معنایی

احسان شریفی[۱]، احمد عبدالله زاده بارفروش[۲*]

*نویسنده مسئول، دریافت: ۱۴۰۰/۰۲/۲۳، بازنگری: ۱۴۰۰/۰۲/۲۷، پذیرش: ۱۴۰۰/۰۳/۰۵

[۱] دانشجوی دکتری، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر، تهران، ایران
[۲] استاد، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر، تهران، ایران

## چکیده

اولین گام برای آغاز یک پژوهش بررسی مطالعه‌های انجام‌شده در حوزه‌ی تخصصی موردنظر است. جمع‌آوری مطالعه‌های مؤثر در حوزه‌ی موردنظر نیازمند یک روش نظام‌مند است. مهندسی نرم‌افزار مبتنی بر شواهد روشی با عنوان مطالعه نگاشتی نظام‌مند را پیشنهاد می‌کند که هدف آن کشف گرایش‌های پژوهشی و دسته‌بندی موضوعات موجود در یک حوزه است. فعالیت اصلی مطالعه نگاشتی نظام‌مند، کلاسه‌بندی مقاله‌ها به‌منظور شناسایی موضوعات پژوهشی موجود در حوزه‌ی موردنظر است. کلاسه‌بندی کلمات کلیدی به‌عنوان یک روش مبتنی بر گراندد تئوری یکی از روش‌های مطرح در این حوزه است. متأسفانه شناسایی کلمات کلیدی مناسب به همراه ارتباطات معنایی بین آن‌ها یکی از چالش‌های مطرح در فرآیند مطالعه نگاشتی نظام‌مند است. در این مقاله یک رویکرد معنایی مبتنی بر آنتولوژی برای غلبه بر چالش فوق معرفی می‌شود. این رویکرد با استخراج مفاهیم کلیدی موجود در حوزه موردنظر فرآیند کلاسه‌بندی را ارتقاء می‌بخشد. به‌منظور بررسی و ارزیابی راهکار فوق یک مطالعه موردی در قالب یک مطالعه نگاشتی نظام‌مند در حوزه‌ی ردیابی ویژگی‌های کیفیتی انجام‌شده است. نتایج ارزیابی نشان می‌دهد که انجام کلاسه‌بندی معنایی با استفاده از مفاهیم کلیدی به دسته‌بندی مدنظر خبرگان حوزه نزدیک‌تر بوده و در شناسایی بهتر و دقیق‌تر موضوعات پژوهشی موجود در یک حوزه نقش به سزایی ایفا می‌کند.

**کلمه‌های کلیدی:** مطالعه نگاشتی نظام‌مند، آنتولوژی، مفهوم کلیدی، کلاسه‌بندی معنایی.

## ۱- مقدمه

برای مرور مطالعه‌های انجام‌شده در یک حوزه پژوهشی، نیازمند استفاده از یک روش جستجوی نظام‌مند[۱] هستیم که امکان شناسایی مطالعه‌های مؤثر انجام‌شده در این حوزه را داشته باشد. «مهندسی نرم‌افزار مبتنی بر شواهد[۲]» دو روش «مطالعه نگاشتی نظام‌مند[۳]» و «مرور ادبیات نظام‌مند[۴]» را معرفی می‌نماید که به‌طور قابل‌اعتماد، قابل‌تکرار، بی‌طرفانه و سخت‌گیرانه امکان انجام این‌گونه مطالعه‌ها را فراهم می‌نمایند[۱]. این دو روش مکمل یکدیگر بوده و هرکدام دارای هدف مشخصی هستند. هدف اصلی مطالعه نگاشتی نظام‌مند که در ادامه با عنوان SMS به آن رجوع می‌کنیم شناسایی و کلاسه‌بندی[۵] گرایش‌های پژوهشی موجود در یک حوزه پژوهشی است. استفاده از این روش هنگام آغاز پژوهش و به‌منظور انتخاب حوزه پژوهشی مناسب است[۲]. هدف از مرور ادبیات نظام‌مند که در ادامه با عنوان SLR به آن رجوع می‌کنیم شناسایی مطالعه‌های اصلی یک حوزه به‌منظور پاسخ‌گویی به یک یا چند سؤال پژوهشی مشخص است. این روش هنگامی‌که از انتخاب حوزه پژوهشی مطمئن شده‌اید مورداستفاده قرار می‌گیرد. سؤال‌های پژوهشی در SMS کلی‌تر بوده و معمولاً فرآیند SMS قبل از SLR انجام می‌شود[۳].

متأسفانه انجام فرآیند SMS به دلیل پیچیدگی نسبی این فرآیند، تعدد فعالیت‌ها، عدم وجود روندهای استاندارد انجام برخی فعالیت‌ها و عدم وجود ابزارهای مناسب با چالش‌هایی همراه است. کلاسه‌بندی گرایش‌های پژوهشی موجود در یک حوزه پژوهشی یکی از فعالیت‌های چالشی است که انجام درست آن تأثیر مستقیمی در کیفیت نهایی فرآیند SMS دارد. به‌منظور انجام کلاسه‌بندی نیازمند استخراج کلمه‌های کلیدی موجود در مقاله‌ها هستیم. از بررسی عنوان، چکیده، متن و کلمه‌های کلیدی تعیین‌شده توسط نویسندگان مقاله یک مجموعه واژگان شناسایی می‌شود که بیانگر مفهوم کلی مقاله است. اگر کلمه‌های کلیدی

صرفاً بر اساس تعداد تکرار انتخاب شوند دقت فرآیند کلاسه‌بندی کاهش می‌یابد. به‌عنوان‌مثال، کلمه‌ی کلیدی «قابلیت ردیابی[۶]» پرتکرارترین کلمه‌ی کلیدی در حوزه‌ی پژوهشی «قابلیت ردیابی نرم‌افزار» است؛ اما این کلمه‌ی کلیدی در فرآیند کلاسه‌بندی تأثیرگذار نیست، زیرا اکثر مقاله‌های این حوزه به‌نوعی مرتبط با این کلمه هستند و نمی‌توان این کلمه را به‌تنهایی به یک خوشه مجزا نسبت داد. مشکل دیگر عدم توجه به رابطه معنایی بین کلمه‌های کلیدی موجود در مقاله‌های مختلف است که باعث می‌شود واژگانی که ازنظر معنایی با یکدیگر شباهت دارند به‌عنوان مشابه شناسایی نشده و از فرآیند کلاسه‌بندی حذف یا در دسته‌های مجزا قرار گیرند. وجود این ایراد باعث می‌شود که هدف اصلی فرآیند SMS که شناسایی گرایش‌های پژوهشی در یک حوزه است با چالش روبه‌رو شده و نتایج کسب‌شده قابل‌اعتماد نباشند.

بدین منظور نیازمند راهکاری هستیم که امکان شناسایی و دسته‌بندی کلمه‌های کلیدی که ازنظر معنایی مشابه هستند را فراهم نماید. در این مقاله و به‌منظور مرتفع نمودن چالش فوق یک رویکرد معنایی بر اساس آنتولوژی معرفی می‌شود. بر اساس این رویکرد یک آنتولوژی در حوزه‌ی موردپژوهش ایجاد می‌شود که امکان شناسایی مفاهیم کلیدی و کلاسه‌بندی آن‌ها را به‌صورت معنایی فراهم می‌کند. به‌کارگیری مفاهیم کلیدی باعث می‌شود که دسته‌بندی موضوعات موجود در یک حوزه‌ی پژوهشی تسهیل شده و باکیفیت و دقت بهتری انجام شود. نتایج ارزیابی رویکرد پیشنهادی بیانگر این است که شناسایی موضوعات پژوهشی در یک حوزه بر اساس رویکرد پیشنهادی به نظر خبرگان حوزه نزدیک‌تر است.

ادامه این مقاله بدین شکل سازمان‌دهی شده است: در بخش دوم پژوهش‌های شاخص انجام شده در این حوزه بررسی می‌شوند. در بخش سوم این مقاله فرآیند SMS برای پژوهش‌گرانی که با این فرآیند آشنا نیستند تشریح می‌شود. در بخش چهارم راهکار پیشنهادی معرفی می‌شود. در بخش پنجم با انجام یک مطالعه موردی نشان می‌دهیم که ارائه این راهکار به همراه معرفی مجموعه‌ای از ابزارها، اجرای فرآیند SMS را تسهیل می‌نماید. در بخش ششم نتایج موردبحث و بررسی قرار می‌گیرند. در بخش هفتم نیز نتیجه‌گیری و کارهای آینده بیان می‌شود.

## ۲- بررسی کارهای انجام‌شده

اولین مطالعه انجام‌شده در خصوص به‌کارگیری دیدگاه مبتنی بر شواهد در مهندسی نرم‌افزار توسط کیچنهام و همکاران[۷] در سال ۲۰۰۴ انجام شد[۱]. در این مقاله رویکرد «پزشکی مبتنی بر شواهد»[۸] به‌منظور استفاده در حوزه مهندسی نرم‌افزار موردبررسی قرارگرفته است. هدف اصلی در این نوع رویکرد دستیابی به اطلاعات دقیق با تحلیل سایر مطالعه‌های اصلی موجود در حوزه‌ی موردنظر است. بررسی مبتنی بر شواهد سعی دارد با بررسی طیف قابل قبولی از مطالعه‌های انجام‌شده در یک حوزه از تعصبات رخ‌داده در نتایج آگاه شده و امکان شناسایی نتایج بی‌طرفانه و قابل‌اعتماد را برای پژوهشگران یک حوزه فراهم کند.

در پژوهش دیگری که توسط دیبا و همکاران[۹] در سال ۲۰۰۵ انجام شد برای اولین بار مهندسی نرم‌افزار مبتنی بر شواهد در قالب یک فرآیند چندمرحله‌ای برای استفاده‌ی پژوهشگران معرفی شد[۴]. نقطه قوت این پژوهش معرفی یک فرآیند برای جمع‌آوری و ارزیابی شواهد است. اما این فرآیند بسیار کلی بیان‌شده و جزئیات و پیچیدگی‌های موجود را به‌صورت شفاف موردبررسی قرار نداده است.

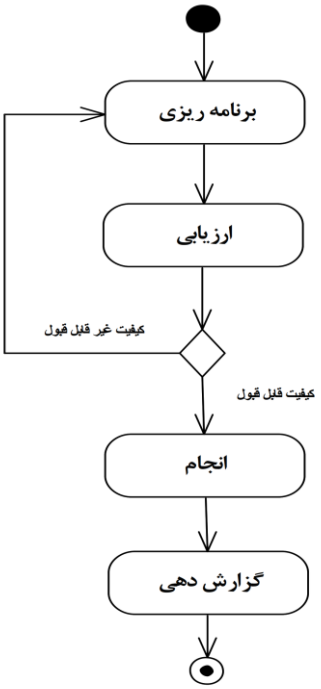
کیچنهام در سال ۲۰۰۶ «مرور ادبیات نظام‌مند» را به‌عنوان یکی از جنبه‌های مهم مهندسی نرم‌افزار مبتنی بر شواهد معرفی و تشریح نمود[۵]. هدف از SLR جمع‌آوری و تلخیص مطالعه‌های پژوهشی مرتبط با یک سؤال پژوهشی مشخص به نحوی است که منصفانه، دقیق و قابل‌کنترل باشد. در این مقاله فرآیند SLR به‌طور دقیق معرفی می‌شود.

پیترسون و همکاران[۱۰] در سال ۲۰۰۸ فرآیند انجام یک مطالعه نگاشتی نظام‌مند را در قالب یک مقاله تشریح نمودند[۶]. در این مقاله علاوه بر معرفی فعالیت‌های مربوط به فرآیند SMS، این فرآیند با SLR مقایسه شده و نحوه انتخاب هر یک بر اساس نیازهای پژوهشگران تشریح شده است.

در سال ۲۰۱۵، پیترسون و همکاران یک راهنمای بروز شده از فرآیند SMS ارائه نمودند که برخی از فعالیت‌های این فرآیند را مورد بهبود قرار داده است[۲]. برای این منظور با انجام یک مطالعه نگاشتی بر روی برخی SMSهای موجود، تکنیک‌ها و روش‌های مورداستفاده در آن‌ها را موردبررسی و مقایسه قرار داده است. نتیجه این پژوهش ارائه یک راهنمای بروز شده برای انجام مطالعه نگاشتی نظام‌مند است که در بخش سوم این مقاله به معرفی آن می‌پردازیم.

## ۳- معرفی فرآیند مطالعه نگاشتی نظام‌مند

فرآیند مطالعه نگاشتی نظام‌مند معرفی‌شده در این مقاله، بر اساس راهنمای[۱۱] ارائه‌شده توسط پیترسون و همکاران است[۲]. شکل ۱ مدل فرآیند SMS بر اساس این راهنما را نمایش می‌دهد. این فرآیند از چهار فاز برنامه‌ریزی، ارزیابی، انجام[۱۲] و گزارش دهی تشکیل شده است. در ادامه فازهای این فرآیند با تمرکز بیشتر بر روی فاز برنامه‌ریزی معرفی می‌شود.




شکل ۱- مدل فرآیند مطالعه نگاشتی نظام‌مند

### ۳-۱- فاز برنامه‌ریزی

فاز برنامه‌ریزی اولین مرحله از فرآیند SMS است. در این فاز کلیه‌ی تصمیم‌های مربوط به انجام SMS اتخاذ می‌شود. جدول‌های ۱ تا ۹ فعالیت‌های فاز برنامه‌ریزی را نمایش می‌دهند. لازم به ذکر است که در هر یک از فعالیت‌های معرفی‌شده در فاز برنامه‌ریزی تکنیک‌ها و روش‌هایی ذکر می‌شود که بررسی جزئیات آن خارج از محدوده‌ی موردنظر این مقاله است. لذا برای درک جزئیات بیشتر به منابع مربوطه ارجاع داده‌شده است. در ادامه فعالیت‌های موجود در این فاز به‌صورت اجمالی معرفی می‌شوند.

جدول ۱- شناسایی نیاز و تعیین محدوده

| | |
|---|---|
| تعیین محدوده مطالعه | کلیه مطالعات انجام شده در حوزه «**ردیابی نیازمندی‌های کیفیتی**» است که در خلال سال‌های **۱۹۹۱** تا **۲۰۱۹** انجام شده است. |
| تعریف سوالات تحقیق | موضوعات پژوهشی اصلی موجود در حوزه ردیابی نیازمندی‌های کیفیتی چیست؟ |

جدول ۲- انتخاب استراتژی جستجو

| جستجوی دستی | جستجو در پایگاه استنادی | گلوله برفی |
|---|---|---|
| ■ | □ | ■ |

جدول ۳- انتخاب استراتژی انجام جستجو

| PICO | انتخاب کلمات کلیدی مقالات شاخص | استانداردها، دایره‌المعارف‌ها و لغت نامه‌ها | مشورت با افراد خبره | بهبود تکراری جستجو برای یافتن مقالات بیشتر |
|---|---|---|---|---|
| □ | ■ | □ | □ | □ |

جدول ۴- تعیین معیارهای شمول و محرومیت

| قواعد تصمیم‌گیری | برطرف کردن اختلاف نظرهای بین پژوهشگران | شناسایی معیارهای عینی و ارزیابی به صورت عینی |
|---|---|---|
| ■ | □ | □ |

جدول ۵- تعیین نحوه‌ی ارزیابی جستجو

| Test-retest | بررسی صفحات وب پژوهشگران شاخص | ارزیابی فرد خبره | مجموعه آزمون |
|---|---|---|---|
| □ | □ | □ | ■ |

جدول ۶- برنامه‌ریزی فرآیند کلاسه‌بندی مستقل از عنوان

| نام نویسنده | سال چاپ | محل چاپ | کشور پژوهش | روش پژوهش |
|---|---|---|---|---|
| ■ | ■ | ■ | ■ | □ |

جدول ۷- برنامه‌ریزی فرآیند کلاسه‌بندی وابسته به عنوان

| دسته‌بندی‌های موجود | کلاسه‌بندی کلمات کلیدی |
|---|---|
| □ | ■ |

جدول ۸- برنامه‌ریزی نحوه بصری‌سازی

| دیاگرام خطی | دیاگرام پای | دیاگرام میله‌ای | دیاگرام حبابی | دیاگرام ون | هیت مپ |
|---|---|---|---|---|---|
| □ | □ | ■ | □ | □ | □ |

جدول ۹- برنامه‌ریزی نحوه بررسی تهدیدات اعتبار

| روایی توصیفی | روایی نظری | روایی تفسیری | تعمیم‌پذیری | تکرارپذیری |
|---|---|---|---|---|
| ■ | □ | □ | □ | ■ |

### ۳-۱-۱- تعیین محدوده مطالعه و تعریف سؤال‌های تحقیق

اولین فعالیت در فاز برنامه‌ریزی، تعیین محدوده مطالعه و تعریف سؤال‌های تحقیقی است که این مطالعه قصد دارد به آن‌ها پاسخ دهد. محدوده‌ی مطالعه بیانگر فاصله‌ی زمانی انتشار مقاله‌های انتخاب‌شده برای انجام مطالعه نگاشتی است.

سؤال‌های پژوهش در دسته‌ی مستقل از عنوان و وابسته به‌عنوان قابل‌طرح هستند. سؤال‌های وابسته به‌عنوان در خصوص شناسایی و کلاسه‌بندی موضوع‌های پژوهشی در حوزه‌ی موردنظر مطرح می‌شود؛ اما سؤال‌های مستقل از عنوان در خصوص اطلاعات کلی مقاله نظیر محل چاپ، نام پژوهش‌گران و نام کنفرانس یا مجله مطرح می‌شود.

### ۳-۱-۲- برنامه‌ریزی نحوه شناسایی مطالعه

برنامه‌ریزی برای فرآیندهای جستجو، انتخاب مقاله‌ها و ارزیابی جستجو در این بخش انجام می‌شود. در ادامه فعالیت‌های موجود در این بخش به‌صورت اجمالی معرفی می‌شود.

#### ۳-۱-۲-۱- برنامه‌ریزی فرآیند جستجو

اولین فعالیت در برنامه‌ریزی فرآیند جستجو انتخاب استراتژی جستجو است. استراتژی جستجو فرآیند انجام جستجو را در قالب تکنیک‌هایی نظیر جستجوی دستی، جستجوی پایگاه‌های استنادی و تکنیک گلوله برفی توصیف می‌کند.

دومین فعالیت در برنامه‌ریزی فرآیند جستجو انتخاب استراتژی انجام جستجو است. هدف اصلی این فعالیت شناسایی رشته و فضای جستجو است که با یکی از تکنیک‌های معرفی‌شده در جدول ۳ انجام می‌شود. فضای جستجو شامل مجله‌ها و کنفرانس‌های مرتبط به حوزه‌ی موردمطالعه است. قبل از شروع جستجو باید مجموعه‌ای از فضاهای جستجو تعریف‌شده و در حین فرآیند به‌صورت افزایشی و تکراری تکمیل شوند. رشته‌ی جستجو عبارتی است که برای جستجوی مطالعه‌های مرتبط در هر یک از نمایه‌های استنادی مورداستفاده قرار می‌گیرد. این رشته بر روی عنوان، چکیده و کل متن مقاله‌های مورد جستجو اعمال می‌شود.

#### ۳-۱-۲-۲- برنامه‌ریزی فرآیند انتخاب مقاله‌ها

در این فعالیت معیارهای شمول[۱۳] و محرومیت[۱۴] تعیین‌شده و استراتژی انتخاب مطالعه تبیین می‌شود. یکی از مهم‌ترین دغدغه‌های هر پژوهش، شناسایی مطالعه‌های باکیفیت انجام‌شده در حوزه موردنظر به‌منظور پاسخگویی به سؤال‌های پژوهشی مطرح‌شده است. لذا باید معیارهای دقیقی برای انتخاب یا حذف مطالعه‌های تعیین شود. هر مطالعه با داشتن حداقل یک معیار محرومیت از لیست مطالعه‌های مشمول نهایی حذف خواهد شد. استراتژی انتخاب مطالعه معمولاً شامل دو مرحله است. در مرحله اول مقاله بر اساس عنوان، کلمه‌های کلیدی، اطلاعات محل چاپ و چکیده مورد ارزیابی قرار می‌گیرد. در مرحله دوم ارزیابی بر اساس کل متن مقاله انجام می‌شود. در هر مرحله درباره شمول یا محرومیت مقاله تصمیم‌گیری می‌شود.

#### ۳-۱-۲-۳- برنامه‌ریزی فرآیند ارزیابی جستجو

چهارمین فعالیت در فاز برنامه‌ریزی، تعیین استراتژی ارزیابی فرآیندهای جستجو و انتخاب مطالعه است. مجموعه آزمون[۱۵] یکی از رایج‌ترین رویکردهای ارزیابی فرآیند SMS است[۲]. بر اساس این روش، یک جستجوی مجزا از طریق مراجعه به پیشگامان[۱۶] این حوزه و بررسی فعالیت‌های پژوهشی آن‌ها آغاز می‌شود. در این فرآیند مجموعه پژوهش‌های مرتبط به حوزه تحقیق جمع‌آوری‌شده و مجموعه آزمون ایجاد می‌شود. به‌منظور ارزیابی فرآیند جستجو مجموعه مقاله‌های نهایی شناسایی‌شده از طریق فرآیند SMS با مجموعه مقاله‌های موجود در مجموعه آزمون مقایسه و ارزیابی نهایی انجام می‌شود. برخی دیگر از روش‌های رایج ارزیابی در جدول ۵ نمایش داده‌شده است.

### ۳-۱-۳- برنامه‌ریزی فرآیند استخراج داده

به‌منظور پاسخگویی به سؤال‌های تحقیق نیازمند استخراج اطلاعات از مقاله‌های مشمول هستیم. انتخاب آیتم‌های داده به نحوی خواهد بود که امکان پاسخگویی به سؤال‌های تحقیق را فراهم نماید. به‌عنوان مثال به‌منظور تعیین مجله‌های فعال در حوزه‌ی پژوهشی موردنظر می‌بایست نام مجله به‌عنوان یک آیتم داده انتخاب‌شده و اطلاعات مربوط به آن از مقاله‌های نهایی استخراج شود. کلمه‌های کلیدی موجود در یک مقاله یکی دیگر از آیتم‌های داده مهمی است که به‌منظور شناسایی موضوع پژوهشی آن مقاله مورداستفاده قرار می‌گیرد.

### ۳-۱-۴- برنامه‌ریزی فرآیند کلاسه‌بندی

مرحله بعدی در این فاز، برنامه‌ریزی فرآیند کلاسه‌بندی است. فرآیند کلاسه‌بندی معمولاً به دو صورت انجام می‌شود. نوع اول «کلاسه‌بندی مستقل از عنوان[۱۷]» است. در این روش مقاله‌ها بر اساس داده‌هایی نظیر نام نویسنده، سال چاپ، محل چاپ که وابسته به محتوا و عنوان مقاله نیستند کلاسه‌بندی می‌شوند. این روش رایج‌ترین روش کلاسه‌بندی مقاله‌ها است که در اکثر مقاله‌های مروری مورداستفاده قرار می‌گیرد.

نوع دوم «کلاسه‌بندی خاص عنوان[۱۸]» است. این نوع کلاسه‌بندی مقاله‌ها را از حیث موضوع و حوزه تحقیق مورد کلاسه‌بندی قرار می‌دهند. درواقع هدف اصلی انجام SMS که شناسایی گرایش‌های پژوهشی موجود در یک حوزه است از طریق این نوع از کلاسه‌بندی پاسخ داده می‌شود. بر اساس راهنمای پیترسون، این نوع از کلاسه‌بندی به دو روش قابل انجام است. اولین روش استفاده از دسته‌بندی‌های موجود است که توسط موسسه‌هایی نظیر IEEE یا SWEBOK انجام‌شده و موردقبول اکثر پژوهش‌گران است. این روش زمانی می‌تواند مورداستفاده قرار گیرد که یک دسته‌بندی استاندارد در حوزه موردنظر موجود باشد.

در روش دوم، بر اساس داده‌های به‌دست‌آمده از فرآیند SMS یک کلاسه‌بندی جدید ایجاد می‌شود. «خوشه‌بندی کلمه‌ی کلیدی[۱۹]» بر اساس این روش است. خوشه‌بندی کلمه‌ی کلیدی مشابه کدگذاری باز[۲۰] در گراندد تئوری[۲۱] است. در روش‌های منطبق بر گراندد تئوری از مجموعه‌ای از داده‌ها به‌صورت استقرایی یک تئوری استخراج می‌گردد. اگر این کار به‌درستی انجام شود بدین معنا است که تئوری به‌دست‌آمده حداقل بر روی یک مجموعه داده[۲۲] به‌طور کامل سازگار است. لذا اگر فرآیند خوشه‌بندی کلمه‌ی کلیدی به‌درستی انجام شود دارای یک‌پایه تئوری است که می‌تواند در هنگام عدم وجود کلاسه‌بندی‌های استاندارد توسط سایر پژوهش‌گران مورد استناد واقع شود.

### ۳-۱-۵- برنامه‌ریزی نحوه بصری‌سازی

تصمیم‌گیری در خصوص نحوه بازنمایی و بصری‌سازی اطلاعات در این بخش اتخاذ می‌شود. جدول ۸ برخی از روش‌های بصری‌سازی اطلاعات را نمایش می‌دهد.

### ۳-۱-۶- برنامه‌ریزی نحوه بررسی تهدید اعتبار

در این مرحله در خصوص نحوه بررسی مهم‌ترین خطراتی که اعتبار یک مقاله مروری نظام‌مند را تهدید می‌کند تصمیم‌گیری می‌شود. روایی توصیفی[۲۳]، روایی نظری[۲۴]، روایی تفسیری[۲۵]، تعمیم‌پذیری[۲۶] و تکرارپذیری[۲۷] پنج نوع اعتبارسنجی معرفی‌شده توسط پیترسون و همکاران است که در جدول ۹ نمایش داده‌شده است[۲].

## ۳-۲- فاز ارزیابی

در فاز دوم از فرآیند SMS فرآیند برنامه‌ریزی مورد ارزیابی قرار می‌گیرد. بدین منظور از متریک‌های ارزیابی ارائه‌شده توسط پیترسون و همکاران استفاده می‌شود[۲]. اگر نتیجه ارزیابی بیانگر کیفیت نامناسب باشد فاز برنامه‌ریزی مورد بازبینی و اصلاح قرار می‌گیرد و این چرخه تا زمانی‌که کیفیت موردنظر حاصل شود ادامه می‌یابد.

## ۳-۳- فاز انجام

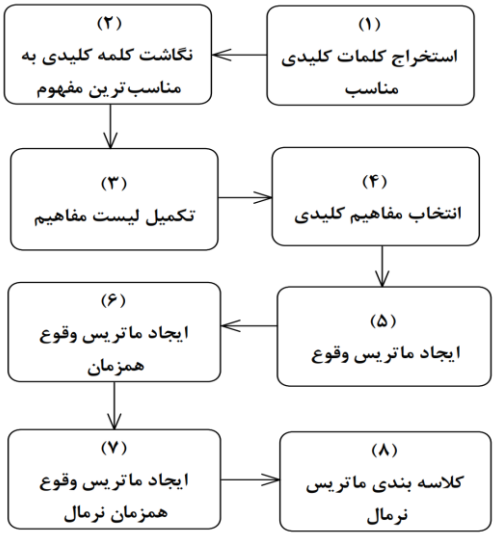
پس از انجام فازهای برنامه‌ریزی و ارزیابی نوبت به اجرای فرآیند SMS می‌رسد. در این فاز فرآیند برنامه‌ریزی‌شده در فاز اول اجرا می‌شود.

## ۳-۴- فاز گزارش‌گیری

نتایج به‌دست‌آمده در فاز انجام در این فاز ساختاردهی شده و قابل‌ارائه می‌شوند. تا جائی‌که امکان دارد ساختار گزارش‌ها باید از یک ساختار نسبتاً یکسان پیروی کنند تا امکان ارزیابی و مقایسه نتایج به‌راحتی فراهم شود.

## ۴- راهکار پیشنهادی

شکل ۲ مدل فرآیند مربوط به راهکار پیشنهادی را نمایش می‌دهد. در ادامه این بخش و در ابتدا ابزارهای مورداستفاده در این فرآیند معرفی می‌شوند. سپس نحوه‌ی ایجاد آنتولوژی حوزه معرفی می‌شود. سپس فعالیت‌های این فرآیند تشریح می‌شوند.




شکل ۲- مدل فرآیند رویکرد پیشنهادی

### ۴-۱- ابزارهای مورداستفاده

انجام هر فرآیند پیچیده بدون استفاده از ابزار مناسب کاری زمان‌بر و پرهزینه است که ممکن است دقت و کیفیت لازم را فراهم نکند. استفاده از یک ابزار مناسب می‌تواند سرعت و دقت انجام فرآیند را افزایش داده و هزینه انجام آن را کاهش دهد. SMS یک فرآیند نسبتاً پیچیده و زمان‌بر است که اگر به‌درستی مدیریت نشود نتایج حاصل از آن قابل‌اتکا نخواهد بود. لذا به‌کارگیری ابزارهای مناسب می‌تواند این فرآیند را تسهیل نموده و از خطاهای انسانی بکاهد. ابزارهای مختلفی درزمینه‌ی انجام SMS و SLR وجود دارد که هر یک دارای مزایا و معایب خاص خود است. جدول ۱۰ لیست ابزارهای مورداستفاده در این پژوهش را نمایش می‌دهد. هر یک ابزارهای معرفی‌شده در بخشی از فرآیند SMS مورداستفاده قرار می‌گیرند. در جدول ۱۲ نحوه‌ی به‌کارگیری این ابزارها در فرآیند SMS بیان‌شده است. به‌غیراز نرم‌افزار اکسل که نرم‌افزاری شناخته‌شده است و موارد کاربرد آن در جدول ۱۲ بیان‌شده است سایر ابزارها در ادامه یک معرفی اجمالی می‌شوند.

پارسیفال[۲۸] یک ابزار آنلاین و متن‌باز[۲۹] است که برای کمک به پژوهش‌گرانی که قصد انجام مطالعه‌های مروری و نگاشتی در زمینه‌ی مهندسی نرم‌افزار را دارند طراحی‌شده است. این ابزار امکان مستندسازی کل فرآیند را فراهم نموده و امکان کنترل مراحل مهم این فرآیند را برای پژوهش‌گر فراهم می‌نماید.

پروتج[۳۰] معروف‌ترین نرم‌افزار طراحی و تولید آنتولوژی است که با فراهم نمودن امکانات متعدد، ساخت آنتولوژی را بسیار آسان نموده است. پروتج قابلیت نمایش گرافیکی آنتولوژی طراحی‌شده را دارد که به درک بهتر ساختار طراحی‌شده کمک بسیار زیادی می‌کند. افزوده شدن موتورهای استنتاجی نظیر پلت[۳۱] و هرمیت[۳۲] امکان بررسی سازگاری یا پایداری آنتولوژی را در نرم‌افزار پروتج فراهم کرده است. همچنین امکان اجرای کوئری با زبانی نظیر اسپارکل[۳۳] نیز در پروتج به‌سادگی امکان‌پذیر است.

بایب‌اکسل[۳۴] ابزاری قوی است که برای انجام تحلیل بر روی‌داده‌های کتابشناسی[۳۵] طراحی‌شده است. این ابزار شامل جعبه‌ابزاری است که امکاناتی نظیر استخراج فیلد، آنالیزهای پایه، تحلیل استناد[۳۶] و محاسبه رخداد هم‌زمان را بر روی فرمت‌های مختلف ارائه می‌نماید. مزیت مهم این ابزار نصب و کاربری آسان آن است.

اس‌پی‌اس‌اس[۳۷] معروف‌ترین نرم‌افزار آماری برای تحلیل داده‌ها به شمار می‌آید. نرم‌افزار اس‌پی‌اس‌اس یکی از قدیمی‌ترین برنامه‌های کاربردی درزمینه‌ی تجزیه‌وتحلیل‌های آماری است که قابلیت انجام آمارهای توصیفی زیبا از اطلاعات، شامل رسم نمودار و چارت‌های گوناگون، محاسبات مربوط به میانگین، انحراف معیار، واریانس و میانه را دارا است. همچنین علاوه بر تحلیل‌های آماری، مدیریت داده‌ها و داده‌کاوی نیز از ویژگی‌های این نرم‌افزار است. یکی از امکانات مهم این نرم‌افزار انجام انواع روش‌های خوشه‌بندی است که در این مقاله به‌منظور انجام خوشه‌بندی سلسله مراتبی از این ابزار استفاده‌شده است.

ابزار کایوگ[۳۸] توسط نویسنده اول این مقاله تولیدشده است. این ابزار با دریافت فهرستی از کلمات کلیدی و بر اساس یک آنتولوژی حوزه، یک لیست توسعه‌یافته از مفاهیم مرتبط به آن را تولید می‌کند.

ابزار آخر بانام آنتولرن[۳۹] نیز توسط نویسنده اول مقاله تولیدشده است. این ابزار مجموعه‌ای از توده متون مرتبط به حوزه‌ی موردنظر را دریافت نموده و بر اساس تکنیک‌های یادگیری یک آنتولوژی حوزه تولید می‌کند.

جدول ۱۰- ابزارهای مورداستفاده در رویکرد پیشنهادی

| نام ابزار | آدرس |
|---|---|
| پارسیفال | https://parsif.al/ |
| پروتج | https://protege.stanford.edu/ |
| بایب‌اکسل | https://homepage.univie.ac.at/juan.gorraiz/bibexcel |
| اکسل | https://products.office.com/en/excel |
| اس‌پی‌اس‌اس | https://www.ibm.com/analytics/spss-statistics- |
| کایوگ | http://islab.ceit.aut.ac.ir/qat-sms |
| آنتولرن | http://islab.ceit.aut.ac.ir/qat-sms |

### ۴-۲- ایجاد آنتولوژی حوزه

بر اساس تعریف گروبر[۴۰]، آنتولوژی یک توصیف صریح و صوری از یک مفهوم‌سازی اشتراکی است[۷]. این تعریف بر دو نکته‌ی کلیدی تمرکز دارد. نخست اینکه آنتولوژی حتماً باید به یک‌زبان صوری که توسط ماشین قابل‌خواندن و پردازش باشد، بازنمایی شود. دوم اینکه مفاهیم و رابطه‌های بین آن‌ها که در یک مفهوم‌سازی مشخص ایجاد می‌شود حتماً باید قابلیت اشتراک داشته باشد و موردتوافق قرار گیرد.

ساخت آنتولوژی به روش‌های دستی و نیمه‌خودکار و خودکار انجام می‌شود. در روش‌های دستی کلیه‌ی مراحل توسعه آنتولوژی به‌صورت دستی و با دخالت فرد خبره انجام می‌شود. این روش‌ها پرهزینه، زمان‌بر و مستعد خطاست. در روش‌های نیمه‌خودکار و خودکار تمام یا برخی از مراحل ساخت آنتولوژی به‌طور خودکار و بدون دخالت انسان انجام می‌شود. خودکارسازی عملیات ساخت آنتولوژی که از آن با عنوان یادگیری آنتولوژی[۴۱] نیز یاد می‌شود هزینه‌ی ساخت را کاهش داده و خطا را کاهش می‌دهد. یکی از نیازمندی‌های اصلی در فرآیند یادگیری آنتولوژی وجود توده متون[۴۲] جامع مرتبط به حوزه‌ی موردنظر است. کیفیت و جامعیت یک آنتولوژی، رابطه مستقیم با جامعیت توده متونی دارد که برای ساخت آنتولوژی از آن استفاده می‌کنیم. خوشبختانه مقاله‌های نهایی انتخاب‌شده در فرآیند SMS یک توده متون باکیفیت برای ایجاد آنتولوژی حوزه فراهم می‌کنند.

در این مقاله به‌منظور تولید آنتولوژی از رویکردی که توسط شریفی و همکاران معرفی‌شده است استفاده می‌کنیم[۸]. هدف از این پژوهش که توسط نویسنده اول این مقاله انجام‌شده است تولید یک آنتولوژی فازی بر اساس رویکرد پیشنهادی توسط لائو و همکاران[۴۳] است[۹]. این رویکرد با ارائه یک ابزار و از طریق تنظیم پارامترها امکان تولید آنتولوژی غیر فازی را نیز فراهم می‌کند.

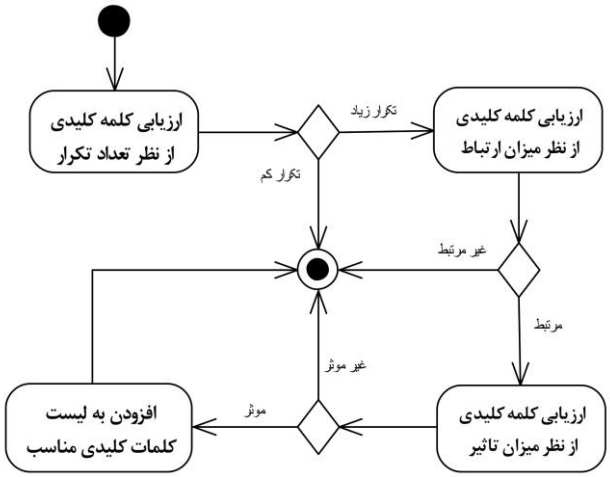
## ۴-۳- شناسایی کلمه‌های کلیدی مناسب

اولین مرحله از راهکار پیشنهادی شناسایی کلمه‌های کلیدی مناسب[۴۴] است. کلمه‌های کلیدی مناسب کلماتی هستند که در تشخیص موضوع مقاله نقش مهمی ایفا می‌کنند. در این بخش طبق فرآیندی که در شکل ۳ مشاهده می‌شود کلمه‌های کلیدی مناسب انتخاب می‌شوند. برای تعیین مناسب بودن یک کلمه کلیدی سه ویژگی تعداد تکرار، مرتبط بودن به حوزه‌ی موردنظر و مؤثر بودن[۴۵] موردبررسی قرار می‌گیرد.

تعیین تعداد تکرار کلمه‌های کلیدی به‌صورت خودکار انجام می‌شود. ابزارهای مختلفی برای استخراج تعداد کلمه‌های کلیدی وجود دارد. در این مقاله از ابزار بایب‌اکسل استفاده‌شده است. یکی از امکانات این ابزار امکان استخراج کلمه‌های پرتکرار در یک متن است. اگر تعداد تکرار یک کلمه‌ی کلیدی از یک حد آستانه بیشتر بود به مرحله‌ی بعد می‌رویم و در غیر این صورت این کلمه حذف می‌شود.

تعیین مرتبط بودن یک کلمه کلیدی به حوزه‌ی موردنظر طی دو مرحله انجام می‌شود. اگر کلمه‌ی کلیدی موردنظر جزء کلمه‌های کلیدی معرفی‌شده توسط نویسنده‌ی مقاله باشد به‌عنوان کلمه‌ی کلیدی مرتبط انتخاب خواهد شد؛ اما اگر کلمه‌ی کلیدی جزء کلمه‌های کلیدی موردنظر نویسنده نباشد ارتباط یا عدم ارتباط آن به حوزه‌ی موردنظر توسط فرد خبره تعیین می‌شود. اگر کلمه‌ی کلیدی موردنظر به‌عنوان مرتبط شناسایی شد به مرحله‌ی بعد می‌رویم و در غیراینصورت کلمه حذف می‌شود.

در آخرین مرحله از این فرآیند میزان مؤثر بودن یک کلمه‌ی کلیدی تعیین می‌شود. بدین منظور یک لیست از کلمه‌های کلیدی غیر مؤثر توسط فرد خبره جمع‌آوری می‌شود. این لیست شامل کلماتی است که تعداد تکرار زیادی دارند و به حوزه مرتبط هستند، اما تأثیری در شناسایی موضوعات پژوهشی ندارند. به‌عنوان‌مثال، «قابلیت ردیابی» پرتکرارترین کلمه‌ی کلیدی در حوزه‌ی «ردیابی نرم‌افزار» است؛ اما این کلمه‌ی کلیدی در فرآیند کلاسه‌بندی تأثیرگذار نیست، زیرا این کلمه‌ی کلیدی در اکثر مقاله‌های این حوزه وجود دارد و نمی‌توان آن را به یک دسته‌ی مجزا نسبت داد. نمونه دیگر «ویژگی‌های کیفیتی» است که یک کلمه‌ی کلیدی غیر مؤثر است. اگر در این مرحله، کلمه‌ی کلیدی موردنظر جزء لیست کلمه‌های کلیدی غیرموثر نبود به لیست کلمه‌های کلیدی مناسب اضافه می‌شود.




شکل ۳- فرآیند انتخاب کلمه‌های کلیدی مناسب

## ۴-۴- استخراج مفاهیم کلیدی

فعالیت‌های دو تا پنج در شکل ۲ بیانگر مراحل استخراج مفاهیم کلیدی است. در این بخش فعالیت شناسایی و استخراج کلمه‌های کلیدی با استفاده از آنتولوژی بهبود می‌یابد. هدف از به‌کارگیری آنتولوژی، کشف ارتباطات معنایی موجود بین کلمه‌های کلیدی و حرکت به سمت ایجاد یک برداشت واحد و اشتراکی از مفاهیم این حوزه به‌منظور شناسایی «مفاهیم کلیدی[۴۶]» است به‌نحوی‌که امکان به اشتراک‌گذاری، همکاری و قابلیت استفاده مجدد این مفاهیم فراهم گردد.

در این بخش هر کلمه‌ی کلیدی مناسبی که در مرحله‌ی ۴-۳ شناسایی‌شده است به یک مفهوم در آنتولوژی که نزدیک‌ترین قرابت معنایی را دارد نگاشت می‌شود. مرحله دوم از فرآیند شکل ۲ بیانگر این فعالیت است. این فرآیند می‌تواند به‌صورت دستی، نیمه‌خودکار و یا خودکار انجام پذیرد. در حالت خودکار میزان شباهت معنایی کلمه‌ی کلیدی مناسب با مفاهیم موجود در آنتولوژی محاسبه‌شده و این کلمه با مشابه‌ترین مفهوم جایگزین می‌شود. در حالت دستی فرآیند نگاشت کلمه‌ی کلیدی مناسب به مفهوم کلیدی مناسب توسط فرد خبره انجام می‌شود. رویکرد این مقاله یک روش ترکیبی است. طبق این رویکرد اگر یک کلمه‌ی کلیدی نتواند توسط محاسبه مشابهت معنایی به یک مفهوم در آنتولوژی نگاشت شود این فرآیند توسط فرد خبره و به‌صورت دستی انجام خواهد شد.

پس از انجام فرآیند نگاشت، لیست مفاهیم کلیدی تکمیل[۴۷] می‌شود. بدین منظور هر مفهوم کلیدی شناسایی‌شده در بخش قبل با مجموعه مفاهیم موجود در سلسله‌مراتب خود تکمیل می‌شود. فرآیند تکمیل مفاهیم بدین شکل است که هر مفهوم کلیدی با فهرستی از مفاهیم پدر خود تا ریشه آنتولوژی توسعه می‌یابد. سپس مفاهیم تکراری از لیست مفاهیم مرتبط با هر مقاله حذف‌شده و لیست مفاهیم توسعه‌یافته ایجاد می‌شود. در انتها با انتخاب مفاهیم پرتکرار از لیست مفاهیم تکمیل‌شده، لیست مفاهیم کلیدی نهایی هر مقاله ایجاد می‌شود. لیست مفاهیم کلیدی کل مقاله‌های مشمول به‌عنوان ورودی فرآیند کلاسه‌بندی استفاده می‌شود.

## ۴-۵- کلاسه‌بندی معنایی

فعالیت‌های ۵ تا ۸ در شکل ۲ بیانگر مراحل کلاسه‌بندی معنایی است. در اولین قدم و با استفاده از لیست مفاهیم کلیدی استخراج‌شده در مرحله قبل ماتریس رخداد[۴۸] ایجاد می‌شود. در تقاطع سطر i و ستون j این ماتریس تعداد تکرار مفهوم کلیدی iام در مقاله jام قرارگرفته است. سپس ماتریس رخداد هم‌زمان[۴۹] ایجاد می‌شود. این ماتریس مربعی بوده و در تقاطع سطر i و ستون j آن، میزان وقوع هم‌زمان دو مفهوم در کنار یکدیگر در کل مقاله‌های مشمول قرار دارد. سپس «ماتریس وقوع هم‌زمان نرمال[۵۰]» از طریق «شباهت کسینوسی[۵۱]» محاسبه می‌شود. در انتها یک الگوریتم خوشه‌بندی روی ماتریس نهایی اعمال می‌شود و دسته‌بندی نهایی استخراج می‌شود.

# ۵- مطالعه موردی

در این بخش یک مطالعه‌ی موردی به‌منظور ارزیابی رویکرد پیشنهادی ارائه‌شده است. بدین منظور یک مطالعه نگاشتی نظام‌مند در حوزه‌ی «ردیابی نیازمندی‌های کیفیتی» انجام‌شده است.

## ۵-۱- انجام جستجو

با توجه به استراتژی جستجوی تعریف‌شده در مرحله اول و بر اساس رشته جستجوی تعیین‌شده تعداد ۳۷۶ مقاله شناسایی شد. مقاله‌های شناسایی‌شده در ابزار پارسیفال بارگذاری شده و فرآیند انتخاب مطالعه آغاز شد. در این مرحله با توجه به فرآیند انتخاب مطالعه، تعداد ۳۶ مقاله به‌عنوان مشمول شناسایی شد. در

ادامه سه مرحله فرآیند گلوله برفی وارونه انجام شد که در مرحله اول ۳۳ مقاله و در مرحله دوم ۱۲ مقاله به‌عنوان مشمول شناسایی گردید. لذا درنهایت از مجموع ۸۷۹ مقاله شناسایی‌شده بر اساس اعمال معیارهای شمول و عدم شمول تعداد ۸۴ مقاله به‌عنوان پذیرفته‌شده نهایی انتخاب شدند.

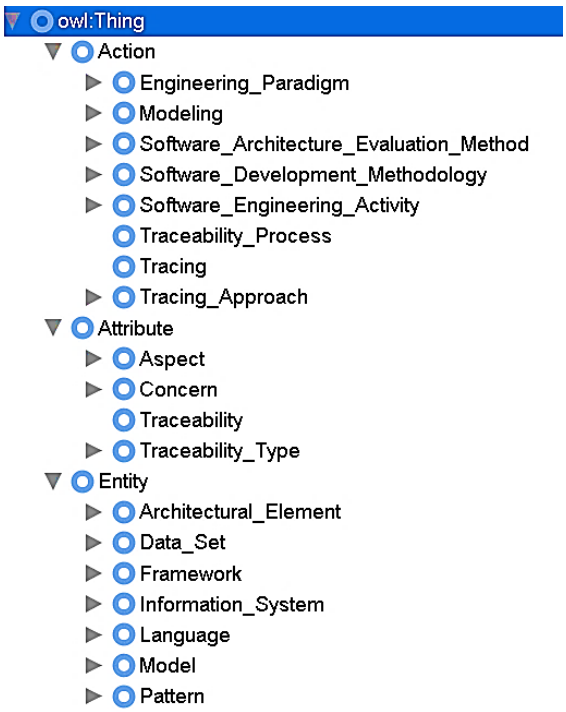
## ۵-۲- ارزیابی فرآیند جستجو

ارزیابی فرآیند جستجو و انتخاب مطالعه در این مقاله با همکاری ۲ نفر از دانشجویان «آزمایشگاه سیستم‌های هوشمند» انجام شد. همان‌گونه که قبلاً بیان شد روش ارزیابی مورداستفاده در این مقاله استفاده از مجموعه آزمون است. مجموعه آزمون موردنظر از طریق مراجعه به مطالعه‌های انجام‌شده توسط پژوهش‌گران شاخص در این حوزه جمع‌آوری می‌شود. نتیجه انجام این مرحله، استخراج ۲۵۴۴ مقاله بود که شامل تمامی مطالعه‌های انجام‌شده توسط پژوهشگران شناسایی‌شده در این حوزه بدون اعمال رشته جستجو بود.

مجموعه مطالعه‌های جمع‌آوری‌شده فوق نیز در ابزار پارسیفال بارگذاری شد و موردبررسی قرار گرفت. از این مجموعه، تعداد ۴۷ مقاله موردپذیرش قرار گرفت که ۷ مقاله در لیست مقاله‌های شناسایی‌شده وجود نداشت. سپس با اعمال فرآیند گلوله برفی بر روی مقاله‌های شناسایی‌شده درنهایت ۵ مقاله به لیست مقاله‌های نهایی افزوده شد و تعداد مقاله‌های نهایی پذیرفته‌شده به ۸۹ عدد رسید. بر اساس نتایج کسب‌شده از این مرحله، میزان شباهت بین مقاله‌های بازیابی شده درروش پیشنهادی بیشتر از ۹۰ درصد بوده که قابل‌قبول است.

## ۵-۳- ایجاد آنتولوژی حوزه

آنتولوژی مورداستفاده در این مقاله یک آنتولوژی سبک‌وزن است که صرفاً بیانگر روابط سلسله مراتبی بین مفاهیم است. لذا سایر فعالیت‌های موجود در فرآیند مفهوم‌سازی نظیر ایجاد روابط دودویی در این مقاله انجام‌نشده است. آنتولوژی مورداستفاده در این مقاله توسط ابزار آنتولرن ایجادشده است[۸]. توده منبع موردنیاز برای ایجاد این آنتولوژی مجموعه مقاله‌های انتخاب‌شده نهایی است. شکل ۴ بخشی از نمودار سلسله‌مراتبی بین مفاهیم را در آنتولوژی حوزه تولیدشده را نمایش می‌دهد.




شکل ۴- بخشی از روابط سلسله‌مراتبی بین مفاهیم در آنتولوژی

## ۵-۴- شناسایی کلمه‌های کلیدی مناسب

فرآیند شناسایی کلمه‌های کلیدی مناسب طبق فرآیند پیشنهادی در شکل ۳ انجام‌شده است. در مرحله اول تعداد ۳۹۷ کلمه‌ی کلیدی بر اساس میزان تکرار شناسایی شد. کلمه‌های با تکرار بیش از ۳ بار در این لیست جای گرفته‌اند. سپس فعالیت بررسی میزان ارتباط کلمه‌ی کلیدی موردنظر به حوزه‌ی موردپژوهش توسط فرد خبره انجام شد و تعداد ۴۰ کلمه‌ی کلیدی غیر مرتبط حذف گردید.

در فعالیت سوم نیز ۱۵ کلمه‌ی کلیدی غیرموثر حذف شد و لیست نهایی کلمه‌های کلیدی مناسب با ۳۴۲ کلمه‌ی کلیدی تشکیل شد. جدول ۱۱ نتایج اعمال فرآیند انتخاب کلمه‌های کلیدی مناسب را بر روی چند کلمه‌ی کلیدی منتخب نمایش می‌دهد.

جدول ۱۱- انتخاب کلمه‌های کلیدی مناسب بر روی چهار کلمه‌ی کلیدی نمونه

| کلمه‌ی کلیدی | تعداد تکرار | مرتبط بودن | مؤثر بودن | مناسب |
|---|---|---|---|---|
| Implicit | زیاد | غیر مرتبط | غیرموثر | ☐ |
| Quality Attribute | زیاد | مرتبط | غیرموثر | ☐ |
| Evolution | زیاد | مرتبط | مؤثر | ■ |
| Architecture | زیاد | مرتبط | مؤثر | ■ |

## ۵-۵- استخراج مفاهیم کلیدی

به‌منظور شناسایی مفاهیم کلیدی، هر یک از کلمه‌های کلیدی مناسب به یک مفهوم در آنتولوژی نگاشت می‌شود. این فرآیند در این مقاله به‌صورت دستی و توسط نویسنده دوم این مقاله به‌عنوان فرد خبره انجام و توسط نویسنده اول مورد بازبینی قرار گرفت. فعالیت تکمیل لیست مفاهیم کلیدی توسط ابزار کایوگ که در بخش ۴-۱ معرفی شد انجام‌شده است. این ابزار لیست مفاهیم مرتبط با هر مقاله را استخراج نموده و این مفاهیم را با استفاده از آنتولوژی توسعه می‌دهد. درنهایت تعداد ۱۹۲ مفهوم کلیدی بر اساس تعداد تکرار انتخاب شد.

## ۵-۶- انجام کلاسه‌بندی معنایی

فرآیند آماده‌سازی داده‌ها برای کلاسه‌بندی و تحلیل توسط نرم‌افزار بایب‌اکسل انجام می‌شود[۱۰]. ماتریس‌های وقوع نیز با استفاده از نرم‌افزار اکسل ایجاد می‌شود. سپس فرآیند کلاسه‌بندی از طریق نرم‌افزار اس‌پی‌اس‌اس اجرا می‌شود. تعداد خوشه‌ها به‌صورت تجربی و با چندین بار اجرای الگوریتم فوق به‌منظور دستیابی به خوشه‌های متوازن تعیین می‌شود.

جدول ۱۲ ورودی و خروجی‌های هر یک از فعالیت‌های شکل ۲ به همراه نوع ابزار استفاده‌شده را نمایش می‌دهد. در ستون اول این جدول شماره فعالیت‌ها بر اساس فرآیند شکل ۲ قرار دارد. در ستون دوم نوع ابزاری که در انجام آن فعالیت استفاده‌شده ارائه‌شده است. در سایر ستون‌ها نیز نوع داده‌های ورودی و خروجی هر فعالیت به همراه اندازه آن‌ها نمایش داده‌شده است.

جدول ۱۲- ورودی‌ها، خروجی‌ها و ابزارهای مورداستفاده در رویکرد پیشنهادی

| ردیف | ابزار | ورودی | خروجی |
|---|---|---|---|
| ۱ | پارسیفال | کلمه‌های کلیدی | کلمه‌های کلیدی مناسب |
| ۲ | پارسیفال | کلمه‌های کلیدی مناسب | مفاهیم کلیدی |
| ۳ | کایوگ | مفاهیم کلیدی | مفاهیم کلیدی تکمیل‌شده |
| ۴ | بایب‌اکسل | مفاهیم کلیدی تکمیل‌شده | مفاهیم کلیدی نهایی |
| ۵ | بایب‌اکسل | مفاهیم کلیدی نهایی | ماتریس رخداد |
| ۶ | اکسل | ماتریس رخداد | ماتریس رخداد همزمان |
| ۷ | اکسل | ماتریس رخداد همزمان | ماتریس رخداد همزمان نرمال |
| ۸ | اس‌پی‌اس‌اس | ماتریس رخداد همزمان نرمال | مفاهیم کلاسه‌بندی شده |

جدول ۱۳ عنوان دسته‌ها و مفاهیم موجود در هر دسته را نمایش می‌دهد. تکنیک ردیابی، ایجاد ردیابی، مدیریت ردیابی، چارچوب ردیابی و الگوریتم ردیابی پنج دسته‌ی شناسایی‌شده بر اساس مقاله‌های مشمول در این حوزه است.

کلیه ابزارهای تولیدشده به همراه ورودی‌ها و خروجی‌های مرتبط با هر یک از فعالیت‌های شکل ۲ بر روی صفحه‌ی مرتبط با آزمایشگاه سیستم‌های هوشمند دانشگاه صنعتی امیرکبیر قابل‌دسترسی است[۵۲].

جدول ۱۳- مفاهیم کلیدی و دسته‌بندی‌ها بر اساس ۵ دسته

| عنوان دسته | مفاهیم کلیدی |
|---|---|
| تکنیک ردیابی | Tracing_Approach, NFR_Tracing_Approach, Custom_Approach, Software_Architecture_Approach |
| ایجاد ردیابی | Software_Engineering_Activity, Trace_Activity, Trace_Creation, Pattern |
| مدیریت ردیابی | Architectural_Element, Trace_Model, Trace_Infrastructure, Trace_Strategy_Model, Architectural_Decision, Model, Trace_Visualization_Model, Trace_Graph |
| چارچوب ردیابی | Framework, NFR_Requirement_Management_Framework, Requirement_Management_Framework, Requirement_Engineering_Activity, Engineering_Paradigm, Requirement_Specification |
| الگوریتم ردیابی | Algorithm, Trace_Retrieval_Algorithm, Trace_Algorithm |

## ۶- بحث

به‌منظور ارزیابی رویکرد پیشنهادی نتایج حاصل از فرآیند «کلاسه‌بندی معنایی» با نتایج حاصل از فرآیند «کلاسه‌بندی کلمه‌ی کلیدی» مورد مقایسه قرار می‌گیرد. بدین منظور فرآیند کلاسه‌بندی سنتی بر روی مجموعه مقاله‌های نهایی اجراشده است. همچنین به‌منظور ارزیابی نتایج حاصل از کلاسه‌بندی، کلمه‌های کلیدی و مفاهیم کلیدی موجود در هر دسته توسط فرد خبره موردبررسی قرارگرفته و صحت قرارگیری آن‌ها در هر دسته ارزیابی می‌شود.

معیارهای مورداستفاده معیارهای استاندارد بازیابی اطلاعات است که شامل دقت[۵۳]، یادآوری[۵۴] و معیار F[۵۵] است.

$$Recall = \frac{a}{a+c} \quad (۱)$$

$$Precision = \frac{a}{a+b} \quad (۲)$$

$$F-measure = \frac{2*Precision*Recall}{Precision+Recall} \quad (۳)$$

پارامترهای a، b و c مورداستفاده در معیارهای فوق در جدول ۱۴ توصیف‌شده است.

جدول ۱۴- توصیف پارامترهای مربوط به روابط ۱ تا ۳

| پارامتر | توصیف پارامتر |
|---|---|
| a | سیستم و فرد خبره با کلمه/مفهوم کلیدی موجود در دسته موافق هستند |
| b | سیستم با کلمه/مفهوم کلیدی موجود در دسته موافق نیست ولی فرد خبره موافق است |
| c | سیستم با کلمه/مفهوم کلیدی موجود در دسته موافق است ولی فرد خبره موافق نیست |
| d | سیستم و فرد خبره با کلمه/مفهوم کلیدی موجود در دسته موافق نیستند |

پارامتر a در فرآیند کلاسه‌بندی بر اساس کلمه‌ی کلیدی بیانگر تعداد کلمه‌ی کلیدی است که توسط فرآیند کلاسه‌بندی در یک دسته قرارگرفته‌اند و درعین‌حال این کلمه‌ها توسط فرد خبره هم به این دسته نسبت داده‌شده‌اند. این پارامتر در فرآیند کلاسه‌بندی بر اساس مفاهیم کلیدی نیز بیانگر تعداد مفاهیمی است که به‌صورت مشترک توسط فرآیند کلاسه‌بندی و فرد خبره در یک دسته جای گرفته‌اند. سایر پارامترهای جدول ۱۴ نیز به همین شکل محاسبه می‌شوند.

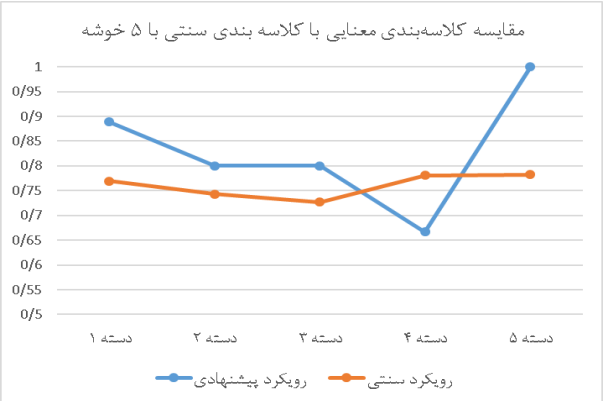
بر اساس مقادیر پارامترها، معیارهای دقت، یادآوری و سنجه‌ی F محاسبه می‌شود. جدول‌های ۱۵ و ۱۶ مقادیر دقت، یادآوری و سنجه‌ی F را برای کلاسه‌بندی بر اساس کلمه‌های کلیدی و مفاهیم کلیدی نمایش می‌دهند. این معیارها در جدول ۱۵ بر اساس انتخاب ۵ خوشه در فرآیند خوشه‌بندی محاسبه‌شده است. بررسی نتایج کلاسه‌بندی مفاهیم کلیدی بیانگر این است که معیارهای دقت، یادآوری و سنجه F در ۴ دسته از ۵ دسته نتایج بهتری نسبت به کلاسه‌بندی کلمه‌های کلیدی ارائه نموده است. بررسی جدول ۱۶ که فرآیند خوشه‌بندی با تعداد ۶ خوشه را نمایش می‌دهد نیز بیانگر این است که تمامی نتایج به‌دست‌آمده در فرآیند کلاسه‌بندی مفاهیم کلیدی بهتر یا معادل نتایج کلاسه‌بندی کلمه‌های کلیدی است. مقادیر سنجه‌ی F در کلاسه‌بندی معنایی و کلاسه‌بندی کلمه کلیدی بر اساس تعداد ۵ و ۶ خوشه در نمودارهای نمایش داده‌شده در شکل‌های ۵ و ۶ مقایسه شده است.

جدول ۱۵- مقایسه نتایج کلاسه‌بندی با ۵ دسته

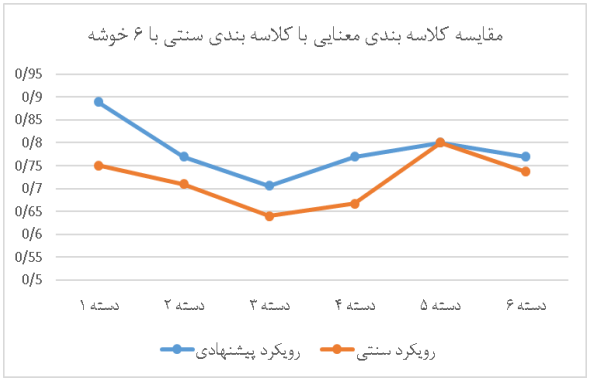
| شماره دسته | کلاسه‌بندی کلمه‌های کلیدی | | | کلاسه‌بندی مفاهیم کلیدی | | |
|---|---|---|---|---|---|---|
| | دقت | یادآوری | سنجه F | دقت | یادآوری | سنجه F |
| دسته ۱ | ۰.۷۱ | ۰.۸۳ | ۰.۷۷ | ۰.۸۰ | ۱ | ۰.۸۹ |
| دسته ۲ | ۰.۷۲ | ۰.۷۶ | ۰.۷۴ | ۰.۶۷ | ۱ | ۰.۸۰ |
| دسته ۳ | ۰.۸۰ | ۰.۶۷ | ۰.۷۳ | ۰.۸۶ | ۰.۷۵ | ۰.۸۰ |
| دسته ۴ | ۰.۸۰ | ۰.۷۶ | ۰.۷۸ | ۱ | ۰.۵ | ۰.۶۷ |
| دسته ۵ | ۰.۷۵ | ۰.۸۲ | ۰.۷۸ | ۱ | ۱ | ۱ |

جدول ۱۶- مقایسه نتایج کلاسه‌بندی با ۶ دسته

| شماره دسته | کلاسه‌بندی کلمه‌های کلیدی | | | کلاسه‌بندی مفاهیم کلیدی | | |
|---|---|---|---|---|---|---|
| | دقت | یادآوری | سنجه F | دقت | یادآوری | سنجه F |
| دسته ۱ | ۰.۸۰ | ۰.۷۱ | ۰.۷۵ | ۰.۸۰ | ۱ | ۰.۸۹ |
| دسته ۲ | ۰.۶۹ | ۰.۷۳ | ۰.۷۱ | ۰.۷۱ | ۰.۸۳ | ۰.۷۷ |
| دسته ۳ | ۰.۷۳ | ۰.۵۷ | ۰.۶۴ | ۰.۷۵ | ۰.۶۷ | ۰.۷۱ |
| دسته ۴ | ۰.۶۹ | ۰.۶۴ | ۰.۶۷ | ۰.۸۳ | ۰.۷۱ | ۰.۷۷ |
| دسته ۵ | ۰.۷۷ | ۰.۸۳ | ۰.۸۰ | ۰.۸۰ | ۰.۸۰ | ۰.۸۰ |
| دسته ۶ | ۰.۷۸ | ۰.۷۰ | ۰.۷۴ | ۰.۸۳ | ۰.۷۱ | ۰.۷۷ |




شکل ۵- مقایسه کلاسه بندی معنایی با کلاسه بندی سنتی با ۵ خوشه




شکل ۶- مقایسه کلاسه بندی معنایی با کلاسه بندی سنتی با ۶ خوشه

## ۷- نتیجه‌گیری و کارهای آینده

در این تحقیق یکی از چالش‌های موجود در فرآیند SMS بررسی و یک راهکار برای بهبود آن پیشنهاد شد. کلاسه‌بندی انجام‌شده در این تحقیق بر اساس گراندد تئوری است. انجام کلاسه‌بندی با این روش اگر به‌درستی انجام شود دارای یک‌پایه‌ی تئوری است که می‌تواند توسط سایر پژوهش‌گران مورد استناد واقع شود. استخراج مفاهیم کلیدی و انجام کلاسه‌بندی معنایی این امکان را فراهم می‌کند که این پایه‌ی تئوری تقویت‌شده و هدف اصلی یک مطالعه نگاشتی نظام‌مند که شناسایی موضوعات پژوهشی در یک حوزه است محقق شود. راهکار پیشنهادی بر پایه‌ی آنتولوژی بوده و امکان شناسایی مفاهیم کلیدی موجود در مقاله‌ها را فراهم می‌کند. استفاده از مفاهیم کلیدی باعث بهبود فرآیند کلاسه‌بندی شده و درنهایت امکان شناسایی دقیق‌تر موضوعات پژوهشی موجود در حوزه‌ی موردنظر را فراهم می‌کند.

نوآوری این مقاله ارائه‌ی رویکردی جدید در حوزه‌ی مطالعه نگاشتی نظام‌مند است که امکان شناسایی موضوعات پژوهشی موجود در یک حوزه را از طریق کلاسه‌بندی معنایی بر مبنای آنتولوژی فراهم می‌کند. نتایج حاصل از این روش بهبود قابل‌توجهی را نسبت به روش سنتی انجام کلاسه‌بندی بر اساس کلمه‌های کلیدی نشان می‌دهد.

اما رویکرد پیشنهادی در این مقاله دارای چالش‌هایی است. ایجاد آنتولوژی یکی از چالش‌های موجود در رویکرد پیشنهادی این مقاله است. اگر در حوزه‌ی موردنظر یک آنتولوژی یا ساختار معنایی از مفاهیم وجود نداشته باشد می‌بایست در یک‌فاز اولیه فرآیند یادگیری آنتولوژی اجراشده و یک ساختار معنایی که دارای روابط رده‌بندی بین مفاهیم باشد ایجاد شود. متأسفانه این فرآیند می‌تواند زمان‌بر و پرهزینه باشد، اما در صورت ایجاد این آنتولوژی امکان به‌کارگیری آن در مطالعه‌های آتی در حوزه موردنظر نظیر انجام SLR و ایجاد درخت تحقیق فراهم می‌شود.

چالش دیگر در رویکرد پیشنهادی یافتن خبرگانی در حوزه‌ی موردنظر به‌منظور انجام برخی از فعالیت‌های رویکرد پیشنهادی است. نحوه‌ی انتخاب افراد خبره، کنترل جانب‌داری[۵۶] در خصوص تصمیم‌های اخذشده توسط آن‌ها و بررسی اعتبار نتایج جزء عواملی هستند که می‌تواند ایجاد مشکل نمایند.

کارهای آتی شامل برطرف نمودن چالش‌های موجود در این مقاله و همچنین بررسی و حل سایر چالش‌های موجود در فرآیند SMS است. تسهیل فرآیند ایجاد آنتولوژی با استفاده از ابزارهای خودکار می‌تواند فرآیند پیشنهادی را بهبود بخشد. همچنین خودکارسازی برخی فعالیت‌ها به‌منظور کم نمودن نقش افراد خبره در رویکرد پیشنهادی می‌تواند جزء کارهای آتی باشد. همچنین در کارهای آتی می‌توان با ارائه‌ی یک چارچوب که علاوه بر فرآیند شامل نحوه انجام فعالیت‌ها و ابزارهای موردنیاز باشد انجام فرآیند SMS برای پژوهشگران حوزه‌ی نرم‌افزار را تسهیل نمود.

## ۸- مراجع


[1] B. A. Kitchenham, T. Dyba, and M. Jorgensen, “Evidence-based software engineering,” in *Proceedings. 26th International Conference on Software Engineering*, May 2004, pp. 273–281, doi: 10.1109/ICSE.2004.1317449.

[2] K. Petersen, S. Vakkalanka, and L. Kuzniarz, “Guidelines for conducting systematic mapping studies in software engineering: An update,” *Information and Software Technology*, vol. 64, pp. 1–18, Aug. 2015, doi: 10.1016/j.infsof.2015.03.007.

[3] B. Kitchenham *et al.* “Systematic literature reviews in software engineering – A tertiary study,” *Information and Software Technology*, vol. 52, no. 8, pp. 792–805, Aug. 2010, doi: 10.1016/j.infsof.2010.03.006.

[4] T. Dyba, B. A. Kitchenham, and M. Jorgensen, “Evidence-based software engineering for practitioners,” *IEEE Software*, vol. 22, no. 1, pp. 58–65, Jan. 2005, doi: 10.1109/MS.2005.6.

[5] B. Kitchenham, “Evidence-Based Software Engineering and Systematic Literature Reviews,” in *Product-Focused Software Process Improvement*, Berlin, Heidelberg, 2006, pp. 3–3, doi: 10.1007/11767718_3.

[6] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, “Systematic mapping studies in software engineering,” in *Proceedings of the 12th international conference on Evaluation and Assessment in Software Engineering*, Swindon, GBR, Jun. 2008, pp. 68–77, Accessed: May 15, 2021. [Online].

[7] T. R. Gruber, “A translation approach to portable ontology specifications,” *Knowledge Acquisition*, vol. 5, no. 2, pp. 199–220, Jun. 1993, doi: 10.1006/knac.1993.1008.

[8] S. Ehsan and D. Mahmood, “AN APPLICABLE METHOD FOR COMPREHENSIVE CORPUS GATHERING NEEDS FOR FUZZY DOMAIN ONTOLOGY GENERATION,” *ELECTRONIC INDUSTRIES*, vol. 7, no. 1, pp. 89–103, Jan. 2016.

[9] R. Y. K. Lau, D. Song, Y. Li, T. C. H. Cheung, and J.-X. Hao, “Toward a Fuzzy Domain Ontology Extraction Method for Adaptive e-Learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 6, pp. 800–813, Jun. 2009, doi: 10.1109/TKDE.2008.137.

[10] F. ASTROM, R. DANELL, B. LARSEN, and J. W. SCHNEIDER, “Celebrating scholarly communication studies,” *A Festschrift for Olle Persson at his 60th Birthday*, vol. 5, 2009.

**احسان شریفی** متولد ۱۳۵۸ است و در حال حاضر دانشجوی دکتری در رشته‌ی مهندسی کامپیوتر گرایش نرم افزار در دانشگاه صنعتی امیرکبیر است. علاقه‌مندی‌های پژوهشی نامبرده شامل کیفیت نرم‌افزار، معماری نرم‌افزار و مهندسی آنتولوژی است. آدرس پست الکترونیکی ایشان عبارت است از:

esharif@aut.ac.ir

**احمد عبدالله‌زاده‌بارفروش** متولد ۱۳۳۲ است. لیسانس حسابداری از دانشگاه تهران، مهندسی کامپیوتر / برنامه‌نویسی از دانشگاه وست‌کوست، دکترا در رشته سیستم‌های هوشمند از دانشگاه بریستول، استاد تمام دانشکده کامپیوتر و فناوری اطلاعات دانشگاه امیر کبیر و مؤسس و مدیر آزمایشگاه سیستم‌های هوشمند است. علاقه‌مندی‌های پژوهشی وی مهندسی نیازمندی، کیفیت نرم‌افزار، معماری نرم‌افزار، مهندسی آنتولوژی و هوش تجاری است. آدرس پست الکترونیکی ایشان عبارت است از:

ahmad@aut.ac.ir

[1] Systematic
[2] Evidence based software engineering
[3] Systematic Mapping Study (SMS)
[4] Systematic Literature Review (SLR)
[5] Classification
[6] Traceability
[7] Kitchenham et al
[8] Evidence-based medicine
[9] Dyba et al
[10] Petersen et al
[11] Guideline
[12] Conducting
[13] Inclusion
[14] Exclusion
[15] Test set
[16] Pioneer
[17] Topic-independent Classification
[18] Topic-specific Classification
[19] Keyword Clustering
[20] Open Coding
[21] Grounded Theory
[22] Data set
[23] Descriptive validity
[24] Theoretical validity
[25] Interpretive validity
[26] Generalizability
[27] Repeatability
[28] Parsifal
[29] Open source
[30] Protege
[31] Pellet
[32] Hermit
[33] Sparql
[34] BibExcel
[35] Bibliographic
[36] Citation
[37] SPSS
[38] CAug
[39] OntoLearn
[40] Gruber
[41] Ontology Learning
[42] Corpus
[43] Lau et al
[44] Suitable
[45] Effectiveness
[46] Key Concepts
[47] Augment
[48] Occurrence Matrix
[49] Co-occurrence Matrix
[50] Normalized Co-occurrence Matrix
[51] Cosine similarity
[52] http://islab.ceit.aut.ac.ir/qat-sms
[53] Precision
[54] Recall
[55] F-measure
[56] Bias

The CSI Journal on Computing Science and Information Technology
Vol. 18, No.2, 2021

# Improvement of the Systematic Mapping Study Process through a Semantic Approach

Ehsan Sharifi[1], Ahmad Abdollahzadeh Barforoush [2]

[1,2] Department of Computer Engineering, Amirkabir University of Technology, Tehran, Iran

## Abstract

The first step in the research is investigating the previous studies of that specialized field. Gathering practical studies in the intended field requires a systematic method. Evidence-based software engineering proposes a technique called Systematic Mapping Study, which aims to discover research trends and classify the available subjects of a field. The main activity of the Systematic Mapping Study is to classify the papers to recognize the available research topics available in the intended field. The keyword classification is based on the Grounded theory and is one of the proposed methods of this field. Unfortunately, recognizing suitable keywords alongside their semantic relationships is one of the main challenges in the Systematic Mapping Study process. In this paper, an ontology-based semantic approach is presenting to overcoming this challenge. This method improves the classification process by extracting the available key concepts of the intended field. A case study in the form of a Systematic Mapping Study was performed in the field of quality attribute traceability to evaluate this solution. The results showed that performing semantic classification using key concepts was closer to the intended classification of the field experts, with an influential role in the better and more accurate recognition of available research topics in a field.

**Keywords**: Systematic Mapping Study; ontology; Key Concept; Semantic Classification