



## دانشگاه صنعتی شیراز

دانشکده مهندسی کامپیوتر و فناوری اطلاعات گروه مهندسی نرم افزار

پایان نامه کارشناسی ارشد

در رشته مهندسی کامپیوتر

طبقه بندی سوالات مطرح شده در فروم های آنلاین با استفاده از استخراج

اطلاعات

نگارش:

عباس نظیری پور

استاد راهنما:

دکتر مرضیه احمد زاده

استاد مشاور:

دکتر رضا اکبری

آبان ماه ۱۳۹۲



## بسمه تعالی

طبقه بندی سوالات مطرح شده در فروم های آنلاین با استفاده از استخراج  
اطلاعات

پایان نامه ارائه شده به عنوان بخشی از فعالیتهای تحصیلی

نگارش:

عباس نظیری پور

برای اخذ درجه کارشناسی ارشد

گروه مهندسی نرم افزار دانشکده مهندسی مهندسی کامپیوتر و فناوری اطلاعات  
دانشگاه صنعتی شیراز

ارزیابی پایان نامه توسط هیات داوران با درجه:.....

دکتر مرضیه احمد زاده استادیار در رشته مهندسی کامپیوتر (استاد راهنما)

دکتر رضا اکبری استادیار در رشته مهندسی کامپیوتر (استاد مشاور)

دکتر رضا جاویدان استادیار در رشته مهندسی کامپیوتر (داور)

---

مدیر امور آموزشی و تحصیلات تکمیلی دانشگاه:

حق چاپ محفوظ و مخصوص به دانشگاه صنعتی شیراز است.

## تأییدیه‌ی صحت و اصالت نتایج

اینجانب عباس نظیری پور دانشجوی رشته مهندسی نرم افزار مقطع تحصیلی کارشناسی ارشد به شماره دانشجویی ۹۰۲۱۴۰۱۰ تأیید می‌نماید کلیه نتایج این پایان‌نامه/رساله، بدون هیچگونه دخل و تصرف، حاصل مستقیم پژوهش صورت گرفته توسط اینجانب است. در مورد اقتباس مستقیم و غیر مستقیم از سایر آثار علمی، اعم از کتاب، مقاله، پایان‌نامه با رعایت امانت و اخلاق علمی، مشخصات کامل منبع مذکور درج شده است.

در صورت اثبات خلاف مندرجات فوق، به تشخیص مقامات ذی صلاح دانشگاه صنعتی شیراز، مطابق قوانین و مقررات مربوط و آئین‌نامه‌های آموزشی، پژوهشی و انضباطی عمل خواهد شد و اینجانب حق هرگونه اعتراض و تجدیدنظر را، نسبت به رأی صادره، از خود ساقط می‌کند. همچنین، هرگونه مسئولیت ناشی از تخلف نسبت به صحت و اصالت نتایج مندرج در پایان‌نامه/رساله در برابر اشخاص ذی نفع (اعم از حقیقی و حقوقی) و مراجع ذی صلاح (اعم از اداری و قضایی) متوجه اینجانب خواهد بود و دانشگاه صنعتی شیراز هیچ‌گونه مسئولیتی در این زمینه نخواهند داشت.

تبصره ۱- کلیه حقوق مادی این اثر متعلق به دانشگاه صنعتی شیراز است.

تبصره ۲- اینجانب تعهد می‌نماید بدون اخذ مجوز از دانشگاه صنعتی شیراز دستاوردهای این پایان‌نامه/رساله را منتشر نکند و یا در اختیار دیگران قرار ندهد.

نام و نام خانوادگی دانشجو:

تاریخ و امضاء

## مجوز بهره‌برداری از پایان‌نامه

کلیه حقوق مادی و معنوی مترتب بر نتایج پایان‌نامه متعلق به دانشگاه و انتشار نتایج نیز تابع مقرارت دانشگاهی است و با موافقت استاد راهنما به شرح زیر، بلامانع است:

- بهره‌برداری از این پایان‌نامه/ رساله برای همگان بلامانع است.
- بهره‌برداری از این پایان‌نامه/ رساله با اخذ مجوز از استاد راهنما، بلامانع است.
- بهره‌برداری از این پایان‌نامه/ رساله تا تاریخ..... ممنوع است.

نام استاد یا اساتید راهنما:

تاریخ:

امضا:

## تقدیرم به

پدر عزیزم:  
او که بهای موفقیت‌م را با جوانیش  
پرداخت.

مادر مهربانم:  
او که بهشت را در برابر محبتش، پیش  
کشی ناپیز از سوی خداوند می‌دانم.

همسر مهربانم:  
که با حضورش، الفبای عشق را برایم معنا  
کرد.

## تشکر و قدردانی:

سپاس خدای را که مرا فرصت داد تا بجویم، بیابم و جرأت‌م داد تا بنگارم. حال که توفیق جمع آوری و تهیه این مجموعه را یافته‌ام بر خود واجب می‌دانم از تمامی عزیزانی که در طی انجام این پژوهش از راهنمایی و یاری‌شان بهره‌مند گشته‌ام تشکر و قدردانی کنم و برای ایشان از درگاه پروردگار مهربان آرزوی سعادت و پیروزی نمایم.

از استاد راهنمای ارجمند سرکار خانم دکتر احمدزاده که با سعه صدر و صبوری مرا راهنمایی نموده و با ارائه نظرات سازنده و رهنمودهای بی‌دریغشان در پیشبرد این پایان‌نامه سعی تمام مبذول داشتند، کمال تشکر را دارم.

از استاد مشاور ارجمند جناب آقای دکتر اکبری که در طول این پژوهش با رهنمودها و تشویق‌های خود مرا مورد لطف خویش قرار دادند، صمیمانه سپاسگزارم.

از داور محترم که زحمت بازخوانی و داوری این مجموعه را به عهده داشتند، صمیمانه تشکر و قدردانی می‌نمایم. همچنین از کلیه اساتید گرانقدر گروه که در دوران تحصیل از محضرشان کسب فیض نمودم، تشکر می‌نمایم.

## چکیده

طبقه بندی سوالات مطرح شده در فروم های آنلاین با استفاده از روش های استخراج

اطلاعات

نگارش:

عباس نظیری پور

در این پژوهش بر روی طبقه بندی سوالات مطرح شده در فروم های آنلاین با استفاده از روش های متن کاوی، کار شده است. با توجه به استقبال روز افزون از انجمن های آنلاین، همچنان مشکلاتی مانند پاسخ دهنده غیر متخصص به سوالات پرسیده شده در انجمن و یا مدت زمان طولانی انتظار برای پاسخ پا برجاست.

انجمن های آنلاینی که هم اکنون در حال استفاده می باشند، قادر به تشخیص مسائلی همچون مخاطب یک سوال در تالارهای گفتگو و یا ارزیابی افراد از طریق عملکرد نوشتاریشان در فروم ها نیستند. در این پژوهش ابتدا ابزارهای استخراج اطلاعات بررسی شده و کاربردی ترین ابزاری که در تجزیه و تحلیل پروفایل افراد موثر است انتخاب گردیده است. این عمل با بررسی دقت نتایج بدست آمده از ابزارهای پر کاربرد در داده کاوی و مقایسه ی آنها با یکدیگر، حاصل شده است. در مرحله ی بعد، با به کارگیری تکنیک های پیش پردازش بر روی مجموعه داده Apple، داده نهایی جهت پردازش بدست آمده است. در این پژوهش مشخص شد که اگر از الگوریتم Kmeans برای خوشه بندی و شبکه های عصبی برای دسته بندی استفاده شود، روش با دقت بالایی حاصل می گردد. بنابراین در فاز پردازش ابتدا با استفاده از الگوریتم داده کاوی K-means و سپس با استفاده از الگوریتم شبکه های عصبی<sup>1</sup> پاسخ دهنده به هر سوال، تخمین زده شد.

**واژه های کلیدی:** متن کاوی ، استخراج اطلاعات، طبقه بندی سوالات، فروم آنلاین،

روش هیبرید، K-means، شبکه های عصبی

---

<sup>1</sup> Neural Network



## فهرست مطالب

۱	<b>۱. فصل اول: کلیات تحقیق</b>
۲	۱-۱- مقدمه
۲	۲-۱- کلیات
۴	۳-۱- هدف پژوهش
۴	۴-۱- اهمیت پژوهش
۵	۵-۱- بیان کلیات مساله
۶	۶-۱- مجموعه داده مورد استفاده
۷	۷-۱- محدودیت های پژوهش
۷	۸-۱- دامنه پژوهش
۸	۹-۱- ساختار پایان نامه
۱۰	<b>۲. فصل دوم: مبانی نظری پژوهش و مروری بر تحقیقات انجام شده</b>
۱۱	۱-۲- مقدمه
۱۱	۲-۲- داده کاوی
۱۱	۱-۲-۲- فنون داده کاوی
۱۳	۲-۲-۲- روش های کشف دانش در داده کاوی
۱۵	۳-۲-۲- معماری یک سیستم داده کاوی
۱۶	۳-۲-۳- متن کاوی
۱۶	۱-۳-۲- متن کاوی چیست؟
۱۸	۲-۳-۲- تفاوت داده کاوی و متن کاوی
۲۰	۳-۳-۲- فرایند متن کاوی
۲۱	۴-۳-۲- روش های متن کاوی
۲۱	۵-۳-۲- کاربردهای متن کاوی
۲۳	۶-۳-۲- وظایف متن کاوی
۲۳	۷-۳-۲- تکنیک های متن کاوی
۲۵	۴-۲- استخراج اطلاعات
۲۶	۱-۴-۲- سیستم های استخراج اطلاعات
۲۷	۲-۴-۲- روش های استخراج اطلاعات
۲۸	۱-۲-۴-۲- روش Text Miner
۳۱	۲-۲-۴-۲- روش مبتنی بر دسته بندی

۳۴	۲-۴-۳- روش های استخراج مبتنی بر قانون آموزش
۳۸	۲-۴-۳- مقایسه روش های استخراج اطلاعات
۴۱	۲-۵-۵- الگوریتم های استفاده شده در این پژوهش
۴۱	۲-۵-۱- الگوریتم K-means
۴۲	۲-۵-۲- الگوریتم شبکه های عصبی
۴۴	۲-۶-۶- معیارهای ارزیابی
۴۷	۲-۷-۷- ابزارهای داده کاوی
۴۸	۲-۷-۱- نرم افزار کلمنتاین
۴۸	۲-۷-۲- نرم افزار وکا
۵۰	۲-۷-۳- نرم افزار رپید ماینر
۵۰	۲-۷-۴- نرم افزار R
۵۱	۲-۸-۸- مروری بر ادبیات پژوهش
۵۱	۲-۸-۱- استخراج اطلاعات
۵۶	۲-۸-۲- تحلیل محتوای انجمن های بحث و گفتگو
۶۱	۲-۹- نتیجه گیری

### ۳. فصل سوم: روش پیشنهادی برای حل مساله

۶۳	
۶۴	۳-۱- مقدمه
۶۴	۳-۲- فرایند پژوهش
۶۸	۳-۲-۱- جمع آوری داده ها
۶۹	۳-۲-۲- پالایش داده ها
۷۰	۳-۲-۳- پیش پردازش داده ها
۷۱	۳-۲-۴- تبدیل متن به بردار
۷۲	۳-۲-۵- میانگین گیری بردار های هر کاربر
۷۳	۳-۲-۶- انتخاب متغیرهای ورودی
۷۵	۳-۳- روش پیشنهادی برای حل مساله / اجرای داده کاوی
۷۹	۳-۳-۱- گام اول: الگوریتم K-means
۸۱	۳-۳-۲- گام دوم: شبکه های عصبی
۸۶	۳-۳-۳- گام سوم: نتایج
۸۶	۳-۴- نتیجه گیری

### ۴. فصل چهارم: نتایج شبیه سازی و تفسیر آنها

۸۸	۱-۴- مقدمه
۸۸	۲-۴- محاسبه معیارهای کارایی
۹۰	۳-۴- ارزیابی روش ارائه شده
۹۱	۴-۴- مقایسه کارایی روش ارائه شده با سیستم پایه
۹۳	۵-۴- نتیجه گیری

## ۹۴ فصل پنجم: خلاصه، نتیجه گیری و ارائه پیشنهادات آتی

۹۵	۱-۵- مقدمه
۹۵	۲-۵- خلاصه
۹۷	۳-۵- پاسخ به سوالات مطرح شده در این پژوهش
۹۸	۴-۵- تحقیقات آینده

۱۰۰ مراجع

۱۰۳ پیوست ها

## فهرست شکل‌ها

- شکل ۱-۱: طرح کلی تحقیق ..... ۹
- شکل ۱-۲: معماری یک سیستم داده کاوی ..... ۱۶
- شکل ۲-۲: فرایند متن کاوی ..... ۲۰
- شکل ۳-۲: چارت ساختاری روش های استخراج اطلاعات ..... ۲۸
- شکل ۴-۲: معماری روش Text Mining ..... ۳۰
- شکل ۵-۲: جریان فرایند استخراج در روش طبقه بندی دو سطحی ..... ۳۴
- شکل ۱-۳: فرایند کلی پژوهش ..... ۶۴
- شکل ۲-۳: فرایند پژوهش ..... ۶۷
- شکل ۳-۳: مرحله پیش پردازش ..... ۷۵
- شکل ۴-۳: چارچوب داده کاوی پیشنهادی ..... ۷۸
- شکل ۵-۳: نمودار مجموع مربعات خطا ..... ۸۰
- شکل ۶-۳: کیفیت خوشه بندی انجام شده ..... ۸۱
- شکل ۷-۳: ساختار یک شبکه عصبی ..... ۸۲
- شکل ۸-۳: جزئیات دنباله های مدل اجرایی ..... ۸۴
- شکل ۹-۳: نمودار اهمیت متغیرهای ورودی در ساخت مدل ..... ۸۵
- شکل ۱۰-۳: چارچوب روش ارائه شده ..... ۸۶
- شکل ۱-۴: چارت بهره ..... ۹۱

## فهرست جدول ها

- جدول ۱-۲: مقایسه داده کاوی و متن کاوی ..... ۱۹
- جدول ۲-۲: مزایا و معایب روش Text Miner ..... ۳۹
- جدول ۳-۲: مزایا و معایب روش مبتنی بر مدل دسته بندی ..... ۳۹
- جدول ۴-۲: مزایا و معایب روش مبتنی بر قانون آموزش ..... ۴۰
- جدول ۵-۲: ماتریس پیچیدگی ..... ۴۵
- جدول ۶-۲: برخی از تحقیقات انجام شده در زمینه استخراج اطلاعات ..... ۵۵
- جدول ۷-۲: برخی از تحقیقات انجام شده در تحلیل محتوای انجمن های بحث و گفتگو ..... ۶۱
- جدول ۱-۳: ویژگی ها در پایگاه داده APPLE ..... ۷۴
- جدول ۱-۴: ماتریس پیچیدگی پایگاه داده APPLE، پیش بینی ها در ستون و هدف های اصلی  
در سطر ..... ۸۹
- جدول ۲-۴: توزیع برچسب ها در کل داده ها ..... ۸۹
- جدول ۳-۴: نتایج ارزیابی برای پایگاه داده APPLE ..... ۹۰
- جدول ۴-۴: مقایسه بین نتایج روش threadCRF و روش ارائه شده در این پژوهش ..... ۹۳

## فهرست نشانه‌های اختصاری

$F_n$	حالاتی که در آن مرجع صحیح و پاسخ غلط است.
$F_p$	حالاتی که در آن مرجع غلط و پاسخ صحیح است.
$K$	تعداد خوشه‌ها
$N$	تعداد کل اسناد
$N_c$	تعداد نمونه‌های هر کلاس
$TF(t,d)$	تعداد دفعات رخداد کلمه $t$ در متن $d$
$T_n$	حالاتی که در آن مرجع و پاسخ هر دو غلط هستند.
$T_p$	حالاتی که در آن مرجع و پاسخ هر دو صحیح هستند.

## فهرست کلمات اختصاری

BIMRF	Binary Markove RandomFerequency
BWI	Boosted Wrapper Induction
IE	Information Extraction
IDF	Inverse Document Frequency
KNN	K-Nearest Neighbors
LSA	Latent Semantic Analysis
MLP	Multilayer Perceptron
NLP	Natural Language Processing
PLSA	Probabilistic Latent Semantic Analysis
POS	Part Of Speach
RBF	Radial Basis Function
SVD	Singular Value Decomposition
SVM	Support Virtual Machine
TF	Term Frequency
VINCA	Visual Intelligent Content Analysis
WEKA	Waikato Environment for Knowledge Analysis

# فصل اول: کلیات تحقیق



## ۱-۱- مقدمه

در این فصل، مسئله‌ی مورد پژوهش، اهمیت آن، محدودیت‌های تحقیق و نیز اهداف مساله معرفی شده است. ابتدا مقدمه‌ای پیرامون موضوع بیان و سپس به کلیات، هدف و اهمیت پژوهش برای آشنایی خوانندگان با موضوع پرداخته شده است، در انتها نیز طرح کلی پایان نامه بررسی خواهد شد.

## ۱-۲- کلیات

در دنیای کنونی مساله مهم کمبود اطلاعات نیست، بلکه کمبود دانشی است که از این اطلاعات می‌توان حاصل کرد. میلیون‌ها صفحه‌ی وب، میلیون‌ها کلمه در کتابخانه‌های دیجیتال و هزاران صفحه اطلاعات در هر شرکت تنها چند دسته از این منابع اطلاعاتی هستند. اما نمی‌توان به طور مشخص، منبعی از دانش را در این بین معرفی کرد. دانش؛ خلاصه‌ی اطلاعات، نتیجه‌گیری، حاصل فکر و تحلیل بر روی آن است.

داده کاوی، یک روش بسیار کارا برای کشف اطلاعات از داده‌های ساخت یافته‌ای که در جداول نگهداری می‌شوند، می‌باشد و همچنین الگوها را از تراکنش‌ها استخراج نموده، داده را گروه بندی و دسته بندی می‌کند. به وسیله‌ی داده کاوی می‌توان به وجود رابطه‌های میان داده‌های پایگاه داده، پی برد.

داده کاوی دارای یک مشکل اساسی است و آن عدم وجود عامیت در کاربرد آن است. تعداد منابع داده‌ای که ساخت یافته، به حد کافی بزرگ و مفاهیم داده کاوی، قابل اعمال بر آنها باشد، چندان زیاد نیستند. در واقع بیشتر دانش ما اگر به صورت غیر دیجیتال نباشند، کاملاً غیر ساخت یافته اند. کتابخانه‌های دیجیتال، اخبار، کتابهای الکترونیکی، بسیاری از مدارک مالی، مقالات علمی و تقریباً هر چیزی که شما می‌توانید در داخل وب بیابید، ساخت یافته نیستند. در نتیجه ما نمی‌توانیم آموزه‌های داده کاوی را در مورد آنها به طور مستقیم استفاده کنیم. بنابراین این مسئله باعث پیدایش

زمینه های علمی جدیدی، مانند تشخیص الگو، بازیابی اطلاعات و استخراج اطلاعات شده است. متن کاوی، داده های متنی غیر ساخت یافته را به منظور استخراج اطلاعات ضمنی پنهان در آن بررسی می کند. بیشتر روش های متن کاوی، الگوریتم ها را بر روی برجسب های مرتبط با هر سند انجام می دهند. این برجسب ممکن است کلمه کلیدی استخراج شده از متن و یا فقط لیستی از کلمات مورد علاقه در متن باشد. به عمل کشف، استخراج و دسته بندی اطلاعات مشابه، استخراج اطلاعات گفته می شود.

استخراج اطلاعات از متن، مزایای بسیاری در برنامه های کاربردی دارد، از جمله: یکپارچه سازی اطلاعات نتیجه شده از وب سایت های مختلف، پاسخ دادن به سوالات، جستجوی اطلاعات تماس، پیدا کردن پروتئین ذکر شده در متن یک مجله ی پزشکی، حذف داده های مشکل دار و یا حتی یافتن و دسته بندی کردن اخبار با موضوع مشخص از بین بسیاری خبرهای موجود در سطح وب است. دسته بندی متن های ساخت یافته به سادگی قابل انجام است در صورتی که جهت دسته بندی متن های بدون ساختار در بین یکسری متون جامع، مثل، اخبار، داده های پزشکی و یا پایگاه های اطلاعاتی، باید نقاط مشترک بین گروه های مشابهی از متن را یافته و سپس متن های مشابه را استخراج و دسته بندی نمود.

این پژوهش روشی برای تجزیه و تحلیل محتوای فروم ها، جهت پیشنهاد بهترین پاسخ دهنده به سوال مطرح شده با استفاده از متن کاوی ارائه می دهد. روش پیشنهادی با استفاده از استخراج پروفایل کاربران و تشخیص تخصص هر فرد در زمینه خاص، سوال مطرح شده را برای او ارسال می کند. در این روش ابتدا محتوای سوال مطرح می شود. در مرحله ی بعد، پاسخ هر یک از کاربران به این سوال، در انجمن آنالیز شده و به بردار تبدیل می شود. سپس با تکرار این فرایند هر یک از کاربران در موضوع های مختلفی که در آن شرکت داشته اند، دارای بردارهای متفاوتی خواهند بود. در ادامه از بردار های هر یک از کاربران به تفکیک موضوع، میانگین گرفته شد، بنابراین هر کاربر در هر موضوعی که در آن شرکت کرده بود دارای تنها یک بردار می شود. در نهایت با مطرح شدن سوالات جدید و مقایسه آن با بردار های میانگین به دست آمده، می توان مشخص کرد که این

سوال به کدامیک از کاربران ارسال گردد.

### ۱-۳-هدف پژوهش

هدف این پژوهش، پیشنهاد بهترین پاسخ دهنده به سوالات مطرح شده، و جلوگیری از سردرگمی در فروم های آنلاین با استفاده از متن کاوی است. در این پژوهش تلاش بر این است که روشی ارائه داده شود که با استخراج پروفایل کاربران، تشخیص تخصص هر فرد و مقایسه سوالات مطرح شده با این پروفایل ها، قادر به تعیین پاسخ دهنده ی این سوالات باشد. جهت رسیدن به این هدف باید روشی پیش بینی کننده ارائه داده شود که هدف این روش، طبقه بندی سوال ها در دسته های متفاوت است. این طبقه بندی می تواند بسیار مفید باشد و با توجه به میزان شباهت هر دسته با پروفایل های مختلف می توان، پاسخ دهنده ای به آن اختصاص داد.

در واقع با این روش، مدیران انجمن (در انجمن های دانشگاهی اساتید) می توانند به طور خودکار طبقه بندی سوالات را انجام دهند. تاکنون در راستای آنالیز محتوای فروم ها، روش های متعددی همانند مارکوف، Kmeans، SVD، مارکوف دودویی و مدل هیستوگرام ارائه شده اند، اما هیچ یک از این روش ها به تنهایی از دقت کافی برخوردار نیستند. در همین راستا، برخی روش های ترکیبی نیز ارائه شده اند، لیکن هدف ما در این پژوهش ارائه روشی ترکیبی است که جهت طبقه بندی سوال های مطرح شده در فروم ها از دقت بالا و پیچیدگی زمانی کم، برخوردار باشد. جهت رسیدن به این اهداف در این پژوهش به سوال زیر پاسخ خواهیم داد:

ترکیب کدام الگوریتم ها روش دقیقی برای مورد مطالعه خواهد داشت؟

### ۱-۴-اهمیت پژوهش

امروزه انجمن های بحث و گفتگو به عنوان مخزن اطلاعات ارزشمند، در حال ظهور می باشند که

اطلاعات، در تعامل بین کاربران انباشته شده است. پاسخ های مطرح شده در هر موضوع، اطلاعات مهمی در مورد محتوای بحث، منتقل می کنند. متأسفانه هیچ یک از انجمن های بحث و گفتگو، قادر به تشخیص مسائلی همچون مخاطب یک سوال در تالار های گفتگو و یا ارزیابی افراد از طریق عملکرد نوشتاری وی در فروم های گفتگو نیستند. در فروم های علمی، این مسائل بسیار مهم اند، و با مشخص بودن مخاطب یک سوال از رد و بدل شدن پیغام های زائد و اشتباه جلوگیری شده و همچنین کاربر زمان کمتری را برای رسیدن به پاسخ صحیح از دست می دهد. بنابراین یکی از مسائل مفیدی که در راستای این پروژه دنبال خواهد شد، طبقه بندی سوالات مطرح شده بر اساس پروفایل هر یک از کاربران، تعیین شخص پاسخ دهنده و ارسال سوال به او می باشد.

## ۱-۵- بیان کلیات مساله

در فروم های آنلاینی که تاکنون در دست می باشند، امکانی برای تشخیص هوشمند پاسخ دهنده به سوالی که در فروم مطرح می شود، با توجه به تخصص او، وجود ندارد. این پژوهش، به دنبال طراحی روشی برای فروم های آنلاین است تا فردی که به عنوان پاسخ دهنده به سوال انتخاب می شود، از طریق میزان معلوماتش در راستای سوال مطرح شده برگزیده شود، و سوالات مطرح شده به افراد خاصی ارسال گردد. گزینش این فرد با استفاده از تکنولوژی استخراج اطلاعات، در پروفایل وی، بحث های انجام شده از طریق او در فروم ها و یا تالارهای گفتگو، موضوع ارتباطات علمیش با سایرین و یا بحث گروه هایی که در آنها شرکت کرده است، انجام خواهد گرفت. با استفاده از این روش می توان، از دریافت پاسخ های ناصحیح، به وجود آمدن افزونگی در ارسال سوال و یا دریافت پاسخ جلوگیری کرد. بنابراین، سر در گمی در یافتن پاسخ صحیح از میان پاسخ های داده شده، کاهش می یابد. و بالاخره می توان با بررسی متن سوال مطرح شده، یافتن کلمات کلیدی خاص و ارتباط آن