

صلى الله عليه وسلم



دانشگاه آزاد اسلامی
واحد علوم و تحقیقات

دانشکده فنی و مهندسی، گروه کامپیوتر

پایان نامه برای دریافت درجه کارشناسی ارشد در رشته مهندسی کامپیوتر (M.S.c)
گرایش: نرم افزار

عنوان:

طراحی نرم افزار کمکیار محقق در ارزیابی مجله مناسب (Web-based)

استاد راهنما:

دکتر سید علی سلیمانی ایوری

استاد مشاور:

دکتر حسین خسروی

نگارش:

رضا میرشاهی

زمستان ۱۳۹۲



معاونت پژوهش و فن آوري
به نام خدا
منشور اخلاق پژوهش

با ياري از خداوند سبحان و اعتقاد به اين كه عالم محضر خداست و همواره ناظر بر اعمال انسان و به منظور پاس داشت مقام بلند دانش و پژوهش و نظر به اهميت جايگاه دانشگاه در اعتلای فرهنگ و تمدن بشري، ما دانشجويان و اعضاء هيئت علمي واحدهاي دانشگاه آزاد اسلامي متعهد مي گرديم اصول زير را در انجام فعاليت هاي پژوهشي مد نظر قرار داده و از آن تخفي نكنيم:

- ۱- اصل حقيقت جويي: تلاش در راستاي پي جويي حقيقت و وفاداري به آن و دوري از هرگونه پنهان سازي حقيقت.
- ۲- اصل رعايت حقوق: التزام به رعايت كامل حقوق پژوهشگران و پژوهيدگان (انسن،حيوان ونبات) و ساير صاحبان حق.
- ۳- اصل مالکيت مادي و معنوي: تعهد به رعايت كامل حقوق مادي و معنوي دانشگاه و کليه همکاران پژوهش.
- ۴- اصل منافع ملي: تعهد به رعايت مصالح ملي و در نظر داشتن پيشبرد و توسعه کشور در کليه مراحل پژوهش.
- ۵- اصل رعايت انصاف و امانت: تعهد به اجتناب از هرگونه جانب داري غير علمي و حفاظت از اموال، تجهيزات و منابع در اختيار.
- ۶- اصل رازداري: تعهد به صيانت از اسرار و اطلاعات محرمانه افراد، سازمان ها و کشور و کليه افراد و نهادهاي مرتبط با تحقيق.
- ۷- اصل احترام: تعهد به رعايت حریم ها و حرمت ها در انجام تحقيقات و رعايت جانب نقد و خودداري از هرگونه حرمت شکني.
- ۸- اصل ترويج: تعهد به رواج دانش و اشاعه نتايج تحقيقات و انتقال آن به همکاران علمي و دانشجويان به غير از مواردی که منع قانوني دارد.
- ۹- اصل برائت: التزام به برائت جويي از هرگونه رفتار غير حرفه اي و اعلام موضع نسبت به کسانی که حوزه علم و پژوهش را به شائبه هاي غير علمي مي آيند.



تعهد نامه اصالت رساله یا پایان نامه

اینجانب رضا میرشاهی دانش آموخته مقطع کارشناسی ارشد ناپیوسته در رشته مهندسی کامپیوتر - نرم افزار که در تاریخ ۱۳۹۲/۱۱/۳۰ از پایان نامه خود تحت عنوان "طراحی نرم افزار کمکیار محقق در ارزیابی مجله مناسب (Web-based)" با کسب نمره ۱۹ و درجه عالی دفاع نموده ام بدینوسیله متعهد می شوم:

(۱) این پایان نامه حاصل تحقیق و پژوهش انجام شده توسط اینجانب بوده و در مواردی که از دستاوردهای علمی و پژوهشی دیگران (اعم از پایان نامه، کتاب، مقاله و ...) استفاده نموده ام، مطابق ضوابط و رویه موجود، نام منبع مورد استفاده و سایر مشخصات آن را در فهرست مربوطه ذکر و درج کرده ام.

(۲) این پایان نامه قبلاً برای دریافت هیچ مدرک تحصیلی (همسطح، پائین تر یا بالاتر) در سایر دانشگاه ها و مؤسسات آموزش عالی ارائه نشده است.

(۳) چنانچه بعد از فراغت از تحصیل، قصد استفاده و هرگونه بهره برداری اعم از چاپ کتاب، ثبت اختراع و ... از این پایان نامه داشته باشم، از حوزه معاونت پژوهشی واحد مجوزهای مربوطه را اخذ نمایم.

(۴) چنانچه در هر مقطع زمانی خلاف موارد فوق ثابت شود، عواقب ناشی از آن را می پذیرم و واحد دانشگاهی

مجاز است با اینجانب مطابق ضوابط و مقررات رفتار نموده و در صورت ابطال مدرک تحصیلی ام هیچ گونه

ادعایی نخواهم داشت.

نام و نام خانوادگی:

رضامیرشاهی

تاریخ و امضاء:

با سپاس فراوان و نهایت قدردانی از :

استاد راهنمای عزیز و گرامی خود، جناب آقای دکتر سید
علی سلیمانی ایوری، به پاس زحمات بی‌وقفه و راهنمایی‌های
گرانقدر ایشان در طی انجام این پژوهش ،

استاد مشاور محترم جناب آقای دکتر حسین خسروی،
بخاطر مساعدت‌های فراوان ،

همچنین اساتید بزرگوارم جناب آقای دکتر زاهدی و جنای
آقای دکتر پویان ،

و در انتها دوستان عزیزم آقایان دکتر قیصر نیکنام و
مهندس حامد رحیم‌اف بخاطر راهنمایی بنده در انجام این
پژوهش .

تقدیم به:

روح پاک پدرم که به من آموخت تا چگونه در
عرصه زندگی مقاومت را تجربه نمایم.

و مادرم، دریای بی‌کران ایثار و فداکاری که وجودم
برایش همه سختی بود و وجودش برایم همه نیکی
است.

و به:

همسر عزیزم که سایه مهربانیش سایه سار زندگیم
می‌باشد، او که اسوه صبر و تحمل و ایستادگی بوده و
مشکلات را برایم آسان نمود.

فهرست مطالب

شماره صفحه

۱

عنوان

چکیده

فصل اول: مقدمه

- ۱- مقدمه ۳
- ۱-۱- پایگاه دانش در سیستم‌های تصمیم‌یار ۵
- ۲-۱- موتور استنتاج در سیستم‌های تصمیم‌یار ۵
- ۳-۱- رابط کاربر در سیستم‌های تصمیم‌یار ۶
- ۴-۱- اهداف پایان‌نامه ۷
- ۵-۱- استنباط آماری و کاربردها ۷
- ۶-۱- منطق فازی و عدم قطعیت ۸
- ۷-۱- داده‌کاوی و کشف دانش ۹
- ۱-۷-۱- داده‌کاوی ۹
- ۲-۷-۱- داده‌کاوی در مقابل آنالیز آماری ۱۰
- ۸-۱- نتیجه‌گیری ۱۱
- ۹-۱- ساختار پایان‌نامه ۱۲

فصل دوم: طراحی و پیاده‌سازی سیستم‌های هوشمند تصمیم‌یار

- ۱-۲- مقدمه ۱۵
- ۲-۲- پیاده‌سازی پایگاه دانش مبتنی بر قوانین ۱۶
- ۱-۲-۲- XML، تکنولوژی برتر جهت پیاده‌سازی قواعد ۱۷
- ۲-۲-۲- تولید و ارائه دانش در تکنولوژی XML ۱۷
- ۳-۲- عملیات متن‌کاوی در پایگاه دانش ۱۹
- ۱-۳-۲- کشف دانش در مقابل متن‌کاوی ۱۹
- ۲-۳-۲- سایر تعاریف متن‌کاوی و KDD ۲۰
- ۳-۳-۲- قسمت‌های جستجوی مربوطه ۲۱

فصل سوم: موتور استنتاج داده‌کاوی پیشنهادی

- ۱-۳- مقدمه ۲۴
- ۲-۳- ریشه‌یابی سند با کلاس Stemmer ۲۶
- ۳-۳- زباله روبی سند با کلاس Word Scraper ۲۶
- ۴-۳- استخراج عناوین موجود در سند توسط کلاس Title Extractor ۲۷
- ۵-۳- توزیع احتمال Co-occurrence در موتور استنتاج داده‌کاوی ۲۷
- ۶-۳- آزمون خی‌دوی پیرسون ۲۹

۳۰	محاسبه نتایج نهایی.....	۷-۳
۳۰	پیاده‌سازی آزمون پیرسون و نتایج نهایی $\chi^2(w)$	۸-۳
۳۴	استخراج دانش بر اساس خلاصه‌سازی استخراجی.....	۹-۳
۳۴	۱-۹-۳ روش فراوانی- عکس فراوانی سند.....	
۳۵	۲-۹-۳ روش مبتنی بر خوشه‌بندی.....	
۳۵	۳-۹-۳ رویکرد نظریه گراف‌ها.....	
۳۶	۴-۹-۳ روش خلاصه‌سازی مبتنی بر منطق فازی.....	
۳۷	۵-۹-۳ روش چکیده.....	
۳۹	۱-۵-۹-۳ مشکلات روش چکیده.....	
۳۹	۱۰-۳ پیاده‌سازی یک مثال عملی در آزمون فرض پیرسون.....	
۳۹	۱-۱۰-۳ پیاده‌سازی کلاس Stemmer در مثال.....	
۳۹	۲-۱۰-۳ پیاده‌سازی کلاس Word Scraper در مثال.....	
۴۰	۳-۱۰-۳ پیاده‌سازی کلاس Title Editor در مثال.....	
۴۰	۴-۱۰-۳ محاسبه توزیع Co-occurrence و وزن کلمات سند در مثال.....	

فصل چهارم: استدلال فازی موتور استنباط CKE

۴۵	۱-۴ مقدمه.....	
۴۵	۲-۴ ساختار یک مجله.....	
۴۷	۳-۴ تطابق تقریبی اسناد با روش‌های فازی.....	
۴۷	۱-۳-۴ روش تطابق تقریبی Dice-Coefficient.....	
۴۹	۲-۳-۴ روش Levenshtein-Distance.....	
۵۰	۱-۲-۳-۴ الگوریتم Levenshtein در یک اپلت جاوا.....	
۵۳	۴-۴ محاسبات نهایی موتور استنباط CKE.....	
۵۵	۵-۴ انتخاب مجلات دارای یک رتبه خاص.....	
۵۵	۶-۴ ارائه مثال‌هایی از پردازش موتور استنتاج CKE.....	

فصل پنجم: نتیجه‌گیری و کارهای آینده

۶۳	۱-۵ نتیجه‌گیری و کارهای آینده.....	
۶۳	۱-۱-۵ نتیجه‌گیری.....	
۶۵	۲-۱-۵ کارهای آینده.....	

فهرست جداول

<u>شماره صفحه</u>	<u>عنوان</u>
۱۱	جدول ۱-۱ - تفاوت‌های داده‌کاوی و آنالیز آماری.....
۳۱	جدول ۱-۳ - متغیرها و پارامترهای توزیع Co-occurrence.....
۳۱	جدول ۲-۳ - میزان تکرار و احتمال کلمات برتر سند.....
۳۲	جدول ۳-۳ - میزان اتفاق همزمان واژه‌های برتر سند با یکدیگر.....
۴۱	جدول ۴-۳ - محاسبات Co-occurrence مربوط به مثال.....
۵۰	جدول ۱-۴ - الگوریتم Levenshtein در یک اپلت جاوا.....
۵۷	جدول ۲-۴ - سه مجله برتر منتخب برای یک مقاله کامپیوتری.....

فهرست نمودارها

<u>عنوان</u>	<u>شماره صفحه</u>
نمودار ۱-۲ - نمایش قواعد If-Then در قالب توابع فازی.....	۱۶
نمودار ۱-۳ - تراکم واژه‌های kind و make با واژه‌های برتر سند.....	۳۳
نمودار ۲-۳ - تراکم واژه‌های imitation و digital computer با واژه‌های برتر س.....	۳۴

فهرست شکل‌ها

<u>شماره صفحه</u>	<u>عنوان</u>
۱۵	شکل ۱-۲ - شمای کلی یک سیستم خبره تصمیم‌یار.....
۲۵	شکل ۲-۲ - نحوه ارتباط یک پایگاه دانش با سایر قسمت‌های یک سیستم پشتیبان تصمیم-گیری.....
۲۵	شکل ۱-۳ - فلوجارت کامل موتور استنتاج پیشنهادی CKE.....
۲۸	شکل ۲-۳ - توزیع آماری کای دو در متلب.....
۳۶	شکل ۳-۳ - گراف میزان تطابق و شباهت اسناد به یکدیگر.....
۳۷	شکل ۴-۳ - دیاگرام مربوط به خلاصه‌سازی فازی سند.....
۳۸	شکل ۵-۳ - دو فاز اصلی روش چکیده.....
۳۸	شکل ۶-۳ - نحوه ادغان در خلاصه‌سازی سند.....
۴۲	شکل ۷-۳ - خروجی نهایی توزیع Co-occurrence مربوط به مثال.....
۴۳	شکل ۸-۳ - مراحل استخراج دانش در موتور استنتاج پیشنهادی CKE.....
۴۶	شکل ۱-۴ - مقایسه‌های فازی قسمت‌های مختلف مقاله و مجله به روش Dice.....
۵۴	شکل ۲-۴ - محاسبات سه بعدی برای هر مقاله.....
۵۷	شکل ۳-۴ - نتایج نهایی پردازش موتور CKE برای یک مقاله کامپیوتری.....
۵۹	شکل ۴-۴ - نتایج نهایی پردازش موتور CKE برای یک مقاله پزشکی.....
۶۱	شکل ۵-۴ - نتایج نهایی پردازش موتور CKE برای یک مقاله کشاورزی.....

چکیده

در دنیای امروز تصمیم‌گیری نقش مهمی را در زندگی بشر ایفا می‌کند. سیستم‌های پشتیبان تصمیم در این زمینه بسیاری از نیازهای بشر را مرتفع می‌سازد. یک نوع سیستم پشتیبان تصمیم که امروزه مورد نیاز است، سیستمی است که بتوان مجله (مجلات) مناسبی را برای مقالات پیشنهاد دهد. محققان و پژوهشگران که دارای مقالات گوناگونی در زمینه‌های مختلف علمی هستند، به دنبال راهکاری برای یافتن مجلات مناسب جهت ارسال^۱ مقالات خود به آنها می‌باشند. داده‌ها و اطلاعات بسیار زیاد موجود و نیز وجود مجلات متعدد در رشته‌های مختلف علمی، باعث سردرگمی پژوهشگران در انتخاب مجله مناسب شده است. بنابراین ارائه یک موتور استنباط هوشمند و دقیق جهت معرفی مجله مناسب برای محققان امری اجتناب‌ناپذیر بوده و طراحی و پیاده‌سازی آن که شامل یک بانک اطلاعات مناسب و روش‌های تصمیم‌گیری نوین مبتنی بر متن‌کاوی و علم فازی است، روش مفیدی برای کمک به کاربران و محققان جهت یافتن مجله (مجلات) مناسب است.

روش پیشنهادی ارائه شده در این پژوهش، طراحی و پیاده‌سازی موتور استنباطی مبتنی بر متن‌کاوی با استفاده از الگوریتم‌های آماری و استدلال فازی است. تکنیک آماری استفاده شده در موتور استنباط معرفی شده در این پژوهش آزمون خی‌دوی پیرسون می‌باشد که آن را توزیع Co-occurrence می‌نامیم. استدلال فازی این موتور استنتاج نیز برگرفته از الگوریتم Dice می‌باشد. موتور استنباط مذکور از سه قسمت اصلی تشکیل شده است: نخست بانک اطلاعات ما که شامل اطلاعات وسیعی از مجلات مناسب در زمینه برق و کامپیوتر با مشخصات دقیق است که این بانک اطلاعات می‌تواند مبتنی بر هر نوع تکنولوژی مانند Xml، Sql Server و یا قواعد Clips باشد. قسمت دوم، الگوریتم‌های تصمیم‌گیری نرم‌افزار می‌باشد که با استفاده از آماری و متن‌کاوی و نتیجه‌گیری فازی طراحی شده است و نتایج مورد نیاز را استخراج و به کاربران نشان می‌دهد. قسمت سوم نیز رابط کاربر^۲ است که بصورت مبتنی بر وب^۳ طراحی شده و قابل انتشار بر روی وب می‌باشد.

واژه‌های کلیدی: تصمیم یار مجلات، پردازش متن در موتور استنتاج، استنباط آماری، موتور استنتاج CKE، بازیابی اطلاعات سند

-
- 1- Submit
 - 2- User Interface
 - 3- Web-based

فصل اول

مقدمه

امروزه انسان با بسیاری از مسائل تصمیم‌گیری روبرو است. دنیای حجیم اطلاعات و واژه‌ها دامنه گسترده‌ای از روش‌ها و رویکردهای تصمیم‌گیری را موجب می‌شود. دانشمندان علوم ریاضی، روش‌های مبتنی بر ریاضیات را کامل‌ترین و بهترین راه رسیدن به تصمیم عاقلانه و صحیح می‌دانند. در مقابل، طراحان روش‌های مبتنی بر ادراک و رویکردهای ادراکی از تطابق بیشتر روش‌های خود با جهان واقع سخن می‌گویند و گروه سومی از مدیران تصمیم‌گیر نیز هستند که نظریه‌های مبتنی بر ریاضیات و نظریه‌های ادراکی را قبول ندارند! این سه گروه دلایل مختلفی را عنوان و اثبات‌هایی را نیز برای حمایت از نظریات خود ارائه می‌کنند.

الگوی صحیح کدام است؟ بهترین روش کدام روش است؟ از سه دسته فوق کدام یک در تصمیم‌گیری موفق‌ترند؟

سیستم‌های پشتیبان تصمیم دوران تکامل و رشد خود را می‌گذرانند و همانند هر سیستم دیگری مجموعه‌ای از طرفداران و منتقدان را به همراه خود دارد. در این میان انسان باید دستاوردهای عینی را تجربه و آینده را نظاره کند و این همان تلاشی است که امروز واقعی را به فردای آرمانی متصل می‌کند. (Califf, 1999).

انسان اگر با فرآیندها و تکنیک‌های تصمیم‌گیری ناآشنا باشد، بی‌شک این اصل را که تصمیم‌گیری بخش اجتناب‌ناپذیری از نقش مدیر و منشاء دید و عملکرد یک مدیر در سازمان است را باور دارد. با این وجود، در میان انبوهی از فعالیت‌های سازمان‌ها و سیستم‌های هوشمندی که برای پشتیبانی از تصمیم‌گیری طراحی شده است، نهایتاً مدیر توانمند مدیری است که به تنهایی تصمیمات بزرگ می‌گیرد! چرا سیستم‌های پشتیبان تصمیم هنوز نتوانسته است جایگاه ویژه‌ای را در زندگی مدیران به خود اختصاص دهد؟ پروفیسور رکس براون، دانشمند بزرگ و استاد تصمیم‌گیری دانشگاه جرج میسون آمریکا که بیش از ۱۸ سال است که در زمینه مشاوره‌های مدیریت فعالیت دارد و تاکنون بیش از ۸۰ کتاب و مقاله در رابطه با تصمیم‌گیری منتشر ساخته است، در یکی از مقالات خود چنین می‌گوید:

"من سالیان دراز است که با مدیران ارشد دولتی کار می‌کنم، با این حال نمی‌توانم ادعا کنم که تعداد زیادی از مشتریان خود را به تصمیم متفاوت و یا بهتری هدایت کرده باشم و فکر نمی‌کنم دیگران هم شرایطی بهتر از این داشته باشند".

وی می‌افزاید: "امکانات تصمیم‌گیری اعم از رویکردهای کمی و یا ادراکی اثرات بسیار محدودی در تصمیم‌گیریها داشته‌اند". امروزه درحالی که بسیاری از مدیران گسترش سیستم‌های پشتیبان تصمیم‌گیری را از مهم‌ترین اولویت‌ها و برنامه‌های سازمانی خود می‌دانند، برخی از صاحب‌نظران با پروفیسور براون موافق و نسبت به سیستم‌های پشتیبان تصمیم اعتماد چندانی ندارند (katedra, 2007). این دو گروه، دلایل و شواهد مختلفی را برای دفاع از نظریات خود ارائه می‌کنند.

سیستم‌های پشتیبان تصمیم در اکثر زمینه‌های مختلف علمی و پژوهشی و صنعتی و سناری کاربرد فراوانی پیدا کرده است. در این نرم‌افزارها کامپیوترها همانند انسان به عمل تصمیم‌گیری می‌پردازند و اعمال انسان را در فرایند تصمیم‌گیری شبیه‌سازی می‌نمایند. به بیان دیگر می‌توان گفت یک سیستم پشتیبان تصمیم مجموعه‌ای از برنامه‌ها و داده‌هایی است که برای کمک به تحلیل و

تصمیم‌گیری طراحی و پیاده‌سازی می‌شوند. نقش کمکی سیستم‌های پشتیبان تصمیم در تصمیم‌گیری بسیار بیشتر از سیستم‌های مدیریت اطلاعات^۱ یا حتی سیستم‌های اطلاعات اجرایی^۲ است.

این سیستم‌ها، یافته‌های انسانی را با الگوریتم‌های کامپیوتری تلفیق می‌کنند تا باعث ارتقاء کیفیت تصمیم‌گیری و اخذ بهترین تصمیم مخصوصاً در مورد مسائل نیمه ساخت‌یافته شوند. سیستم‌های پشتیبان تصمیم دارای یک بانک اطلاعاتی متشکل از دانش مربوط به یک موضوع و یک الگوریتم تصمیم‌گیری پیچیده و یک سیستم مدلسازی برای آزمایش تصمیمات بدست آمده هستند. هدف هر سیستم پشتیبان تصمیم‌گیری دستیابی به سه مورد زیر است:

- یاری رساندن به مدیران در پروسه تصمیم‌گیری در مسائل نیمه ساخت‌یافته
- کمک به تصمیم‌گیری انجام شده توسط مدیر و نه جایگزینی سیستم به جای مدیر
- افزایش کارایی، دقت و صحت تصمیم‌گیری در سازمان‌ها.

هر سیستم پشتیبان تصمیم از سه قسمت اصلی زیر تشکیل شده است:

- پایگاه دانش^۳
- موتور استنباط^۴
- رابط کاربر

۱-۱-

پایگاه دانش در سیستم‌های تصمیم‌یار

پایگاه دانش حاوی اطلاعات بسیار مهم و دانش هر سیستم خبره است. در این پایگاه، دانش خبره به صورت قابل فهم برای سیستم ذخیره می‌شود. در اینجا باید دو اصطلاح زیر را تعریف کنیم:

- شیء^۵: منظور از شیء موجودیتی مستقل است که دارای صفات و ویژگی‌های خاص خود می‌باشد.
- خصوصیات^۶: منظور از خصوصیت یا «صفت»، یک ویژگی است که با توجه به قوانین موجود در آن، شیء را توصیف می‌کند و در واقع شیء را از زوایای مختلف وصف می‌کند. در واقع می‌توان پایگاه دانش خود را به صورت لیستی از اشیاء گوناگون که در آن خصوصیات و شاخص‌های مربوط به هر شیء در ادامه آن شیء آمده باشد، در نظر گرفت (Califf, 1999). در کوتاهترین و ساده‌ترین حالت (که در اکثر کاربردها این حالت مد نظر است) قانونی که به یک خصوصیت اعمال می‌شود، بیان‌کننده این موضوع است که آیا شیء مورد نظر ما یافت شده است یا خیر؟ معمولاً پایگاه دانش به صورت مجموعه‌ای از قوانین^۷ پیاده و ذخیره‌سازی می‌شود که پرکاربردترین آن‌ها قوانین If-Then می‌باشد.

-
- 2- Management Information Systems (MIS)
 - 3- Executive Information Systems (EIS)
 - 1- Knowledge Base
 - 2- Inference Engine
 - 3- Object
 - 4- Attributes
 - 5- Rules

پس از پیاده‌سازی پایگاه دانش و اشیاء مربوطه و خصوصیات توصیف کننده مربوط به اشیاء، سیستم تصمیم‌یار باید مشخص کند که کدام قوانین را برای حل مسئله خاصی به کار می‌برد. همچنین سیستم باید مشخص کند که این قوانین را در کدام زمینه و متناسب با کدام موضوع به کار می‌برد.

موتور استنتاج در حقیقت قلب تپنده هر سیستم خبره است. موتور استنتاج یک سیستم پیشرفته است که قوانین استنتاج را که به صورت مجموعه‌ای از "اگر ... آنگاه ..." می‌باشد، برای یافتن پاسخ نهایی به کار می‌گیرد. ماهیت اصلی سیستم خبره به الگوریتم‌ها و روش‌هایی است که این قوانین بر اساس آن‌ها پردازش می‌شوند. موتور استنتاج برای رسیدن به نتیجه نهایی می‌تواند به دو روش عمل کند و در واقع سلسله مراتب قواعد استدلال را به دو طریق پردازش کند. روش اول، روش استدلال پیش‌رو است که از داده‌ها شروع کرده و به نتیجه مطلوب خود می‌رسد. در واقع از (اگر) ها شروع کرده و به نتایج یا (آنگاه) های مناسب می‌رسد. به عبارت دیگر در روش زنجیره پیش‌رو از تالی به مقدم‌ها می‌رسیم. روش دوم استنتاج در نرم افزار تصمیم‌یار آن است که از نتایج شروع می‌کند و به دنبال مقدمات (تالی‌ها) با شرایط صحیح و مناسب اولیه می‌گردد. به بیان ساده نقطه شروع (آنگاه) ها هستند و از آنگاه‌ها به (اگر) ها دست می‌یابد. روش اول استنتاج را روش مبتنی بر مقدم و روش دوم را روش مبتنی بر تالی یا روش مبتنی بر هدف می‌گوییم.

۱-۳- رابط کاربر در سیستم‌های تصمیم‌یار

رابط کاربر که می‌تواند به صورت مبتنی بر سیستم‌عامل^۱ و یا مبتنی بر وب طراحی و برنامه‌نویسی شود، لایه‌ای بین موتور استنتاج و کاربر است. توسط این لایه کاربران داده‌های خود را به سیستم هوشمند داده و اطلاعات بدست آمده را از سیستم دریافت می‌نمایند. رابط کاربر از نظر زیبایی و طراحی گرافیکی بسیار حائز اهمیت بوده و کاربر را به سیستم نزدیکتر می‌کند. معمولاً از یکی از زبانهای برنامه‌نویسی بصری و یا مبتنی بر وب جهت پیاده‌سازی رابط کاربر استفاده می‌شود.

چند نمونه از سیستم‌های پشتیبان تصمیم عبارتند از:

- پزشکیار
- سیستم پشتیبان تصمیم بالینی
- تصمیم‌یار روانشناسی سپهر
- سیستم‌های پشتیبان تصمیم GIS

سیستم‌های پشتیبان تصمیم فوق از موتور استنتاج فازی بهره می‌برند و هر کدام دارای درصد خطایی می‌باشند. این درصد خطا نشان‌دهنده این موضوع است که این‌گونه سیستم‌ها هنوز به مرحله نهایی تکامل خود نرسیده‌اند و تحقیقات بسیار لازم است تا دقت این سیستم‌ها به حد مطلوب برسد. نکته دیگر این که موتور استنتاج هیچ‌کدام از سیستم‌های فوق از الگوریتم‌های متن‌کاوی و داده‌کاوی استفاده نمی‌کند و فقط از قوانین و استنتاج فازی بهره می‌برد. تفاوت کار ما در این است که موتور استنتاج پیشنهادی ما از متن‌کاوی استفاده کرده و در انتها برای تجزیه و تحلیل نهایی مجموعه جواب‌ها از یک روش فازی استفاده می‌کند.

1- OS-Base

در این پایان‌نامه هدف ما طراحی و پیاده‌سازی نرم‌افزاری هوشمند می‌باشد که با تکیه بر الگوریتم‌های آماری و متدهای متن‌کاوی و استخراج دانش از متن، سعی دارد تا اطلاعات مقاله یا مقالات کاربر (عنوان مقاله، چکیده، کلمات کلیدی و ...) را گرفته و سپس با بانک اطلاعات خود که حاوی مجلات برق و کامپیوتر می‌باشد، تطابق داده و مجلات مناسب را برای ارسال مقاله کاربر جستجو کرده و به کاربر نمایش دهد. شایان ذکر است که مجلات یافت شده را با میزان تطابق بدست آمده از روشهای فازی، به کاربر معرفی خواهد کرد. خروجی نرم افزار، اطلاعات و داده‌های فازی است که غیرفازی شده‌اند و باید حداقل خطای ممکن را داشته باشد. طراحی و پیاده‌سازی نهایی به صورت مبتنی بر وب می‌باشد که امروزه بسیار رایج بوده و قابلیت پیاده‌سازی در دنیای اینترنت را دارد و نیز بر روی تمام سیستم‌عامل‌ها قابل اجراست.

۱-۵- استنباط آماری و کاربردها

امروزه علم آمار استنباطی در اکثر تحقیقات علمی و پژوهشی کاربرد فراوانی پیدا کرده است. آماری که به استنباط و تصمیم‌گیری می‌پردازد. این زمینه دارای ابزارها و آزمون‌هایی است. مثلاً آزمون t ، آزمون نرمال و ضرورت و اهمیت این ابزارها و آزمون‌ها را می‌توان در تصمیم‌گیری، تجزیه و تحلیل و توصیف داده‌های حاصل و گسترش یافته‌های بدست آمده، مشاهده کرد. بنابراین در عمل، روش‌های آماری چندین نقش ایفا می‌کنند که با هم ارتباط دارند. متدهای آماری برای خلاصه کردن و توصیف داده‌ها راهکارهای لازم را فراهم می‌سازند.

علم آمار بر دو نوع توصیفی و استنباطی است. وظیفه آمار توصیفی جمع‌آوری، خلاصه کردن، بیان و تحلیل اطلاعات کمی مستخرج از نمونه‌ها یا جامعه‌ها است. اما در استنباط و تصمیم‌گیری‌ها معمولاً کار با توصیف اطلاعات و داده‌ها و تشریح آن‌ها تمام نمی‌شود، بلکه آنچه را که از توصیف و بررسی جامعه کوچک یا نمونه به دست آورده‌ایم به جوامع مشابه بزرگتر بسط و گسترش می‌دهیم. در اکثر مواقع، مطالعه تمام اعضای یک جامعه و شناخت کامل جامعه غیرممکن است. بنابراین یک پژوهشگر به روش‌هایی محتاج است که بتواند با استفاده از آنها نتایج به دست آمده از بررسی گروه‌های کوچک را به گروه‌های بزرگتر بسط و گسترش دهد. در حقیقت به مجموعه این عملیات آمار استنباطی گوئیم. یعنی استنباط و استدلال بر روی یک جامعه بزرگ از روی تفسیر یک جامعه کوچک. به زبان ساده، در اکثر پژوهش‌ها و علوم مختلف کسب اطلاعات درباره جوامع کوچک هدف نهایی پژوهشگر نیست، بلکه هدف او این است که از طریق یافته‌های این گروه کوچک، جامعه بزرگتر را تفسیر کند و اطلاعات لازم را درباره جامعه بزرگی که این گروه کوچک مربوط به آن است، کسب کند. یعنی در این پژوهش‌ها هدف محقق تعمیم نتایج به دست آمده از یک جامعه کوچک به یک جامعه بزرگتر می‌باشد. این تعمیم و گسترش مستلزم آن است که از علم خاصی از آمار تحت عنوان "استنباط آماری" استفاده نماید (Raymond, 2005).

۱-۶- منطق فازی^۱ و عدم قطعیت

منطق فازی و نظریه عدم قطعیت و ابهام در علم برای اولین بار توسط پرفسور لطفی-زاده، استاد علوم کامپیوتر دانشگاه برکلی کالیفرنیا^۲ بیان شد.

1- Fuzzy Logic

2- California Berkeley university

علم فازی ابداع شده توسط لطفی زاده نگرشی جدید به ابهام در علم است. مقاله‌ای از لطفی زاده تحت عنوان "شرحی بر دیدی نو در تجزیه و تحلیل سیستم‌های پیچیده و فرایندهای تصمیم‌گیری" مفهوم تام استفاده از متغیرهای زبانی را در سیستم‌های حافظه و کنترل بیان کرد. این مقاله مبنای علم کنترل بر اساس منطق فازی است که در آینده تاثیر بیشماری در طراحی سیستم‌های کنترل هوشمند خواهد داشت. در ابتدا منطق فازی فقط در محیط دانشگاه زنده بود و بیرون از دانشگاه اهمیتی به آن داده نمی‌شد. با اینکه امروزه منطق فازی بسیار فراگیر و پرکاربرد شده است ولی لطفی زاده منطق فازی را تمام ماجرا نمی‌داند. وی می‌گوید: "کارهای زیادی هست که انسان می‌تواند به آسانی انجام دهد در حالی که کامپیوترها و سیستم‌های منطقی فازی یک فرا مجموعه از منطق بولی است که بر مفهوم درستی نسبی دلالت می‌کند" (Dubois, 2005). منطق کلاسیک یا ارسطویی، جهان واقع را بر اساس یک روش دوتائی نشان می‌دهد (درست یا غلط، ۰ یا ۱، سیاه یا سفید)، ولی منطق فازی صحت هر چیزی را با یک عدد که مقدار آن بین صفر و یک متغیر است، بیان می‌کند. مثلاً اگر سیاهی را عدد صفر و سفیدی را عدد ۱ بدانیم، رنگ خاکستری عددی بین صفر و یک خواهد بود.

لطفی زاده در نظریه‌های فازی مطرح شده‌اش معتقد است که وجود ابهام در ماهیت علوم مختلف است. بر خلاف برخی محققان که معتقدند باید دقت تقریب‌ها را بیشتر کرد تا بهر موری بیشینه شود، لطفی زاده معتقد است ابهام را باید به عنوان بخشی از علم قبول کرد. در منطق ارسطویی، همه چیز به شکل درست یا نادرست چیده می‌شود. تمام واقعیات یا درست یا نادرست هستند. بنابراین جمله «هوا گرم است»، در مدل ارسطویی یک واقعیت نمی‌باشد، زیرا میزان گرم بودن برای هر انسانی و در هر مکانی متفاوت است و این جمله مطلقاً همیشه درست یا همیشه نادرست نیست. بر اساس منطق فازی، عبارات و جملاتی هستند که مقداری درست و مقداری نادرست هستند که این مقدار با استفاده از اعداد بیان می‌شود.

منطق فازی دارای ابعاد گسترده‌ای است که در اینجا مجال بیشتر برای این بحث نیست و در فصل چهارم این پژوهش روشهایی از آن را در عمل پیاده‌سازی می‌کنیم.

داده‌کاوی و کشف دانش

۷-۱-

۱-۷-۱- داده‌کاوی

امروزه با حجم وسیع اطلاعات و داده‌ها به تکنیک‌ها و راهکارهایی نیازمندیم تا پس از پردازش داده‌ها، اطلاعات اصلی و معنادار که حاوی دانش مهمی است را از این داده‌های حجیم بیرون کشیده و در موارد مختلف مورد استفاده قرار دهد.

با گسترش رهیافت پایگاه‌های اطلاعات متعدد در زمینه‌های گوناگون و حجم بالای داده‌های ذخیره شده در این پایگاه‌ها، به الگوریتم‌هایی نیازمندیم تا با استفاده از آن بتوان این داده‌ها را پردازش و تجزیه و تحلیل و اطلاعات حاصل از آن را استخراج و اختیار کاربران قرار داد (Mannila, 2005).

یکی از روشهای بسیار مهم برای رسیدن به هدف فوق که توسط آن می‌توان الگوهای مفیدی را در میان داده‌ها بدست آورد و بوسیله آن الگوها، اطلاعات مهمی را استخراج کرد، داده‌کاوی است. این علم که جزء علوم نوپای کامپیوتر محسوب می‌شود، اطلاعاتی را در اختیار مدیران و تحلیل‌گران قرار می‌دهد تا براساس آنها تصمیمات مهم و حیاتی را اتخاذ نمایند.

شایان ذکر است که اصطلاح داده‌کاوی در مواقعی کاربرد دارد که با حجم وسیعی از داده‌ها، در حد مگا یا ترابایت روبرو باشیم. در عمل داده‌کاوی در اسناد و پایگاه داده، هرچه داده‌ها بیشتر باشد و وابستگی میان داده‌ها گسترده‌تر و پیچیده‌تر باشد، عمل داده‌کاوی مشکل‌تر و بدست آوردن الگوها سخت‌تر است. اصلی‌ترین عاملی که سبب شده است داده‌کاوی در صنعت بسیار مورد استفاده قرار گیرد، وجود اطلاعات وسیع در صنعت است (نظیر اطلاعات فنی تمام خودروهای جهان). (Raymond, 2005).

داده‌کاوی از چندین زمینه علمی بطور توأم استفاده می‌کند و منحصر به زمینه خاصی از علوم نیست، مانند: تکنولوژی پایگاه داده، هوش مصنوعی، شبکه‌های عصبی، آماری، سیستم‌های مبتنی بر دانش و

مطالب گفته شده را می‌توان از نگاه تاریخی به اختصار به شش مرحله تاریخی زیر تقسیم کرد:

- مرحله اولیه: پیدایش پایگاه اطلاعات و جمع‌آوری داده‌ها (تا دهه ۱۹۶۰)
- مرحله دوم: سیستم‌های مدیریتی متشکل از پایگاه اطلاعات (دهه ۱۹۷۰ و اوایل دهه ۱۹۸۰)
- مرحله سوم: سیستم‌های پایگاه اطلاعاتی پیشرفته (اواسط دهه ۱۹۸۰ تا زمان حاضر)
- مرحله چهارم: داده‌کاوی در اطلاعات ذخیره شده در پایگاه اطلاعات (اواخر دهه ۱۹۸۰ تا به امروز)
- مرحله پنجم: تکنولوژی پایگاه اطلاعاتی شبکه‌ای (دهه ۱۹۹۰ تاکنون)
- مرحله ششم: تحول اساسی در پایگاه اطلاعاتی و انسجام آن (از ۲۰۰۰ به بعد)

۱-۷-۲- داده‌کاوی در مقابل آنالیز آماری

علم و راهکار داده‌کاوی (که گاهی اوقات کشف اطلاعات یا کشف دانش نامیده می‌شود) عبارت است از تکنیک‌ها و روش‌هایی که از زوایای مختلف به تجزیه و تحلیل داده‌ها پرداخته و پس از تجزیه و تحلیل، الگوهای مفیدی را استخراج می‌کند. این الگوهای کشف شده را می‌توان برای اهداف و مقاصد مختلف از جمله مقایسه و کاربرد، تصمیم‌گیری و یا هر دو به کار برد. نرم افزار داده‌کاوی ابزار تجزیه و تحلیل برای اطلاعات مختلف است (Feldman, 2003). این نرم افزار به کاربران امکان می‌دهد اطلاعات را از ابعاد و زوایای متفاوت تحلیل و طبقه‌بندی کنند که این عمل را خوشه‌بندی می‌نامند.

به بیان علمی، داده‌کاوی عبارت از فرآیندی است که در میان اطلاعات وسیع و داده‌های بزرگ از جمله بانک‌های اطلاعاتی بزرگ و گسترده روابط و الگوهای را پیدا می‌کند که این روابط منجر به استخراج اطلاعات و دانش مفید می‌گردد. این نکته حائز اهمیت است که داده‌کاوی و آنالیز آماری دو مقوله جداست. در جدول ۱-۱ تفاوت این دو آورده شده است:

جدول ۱-۱- تفاوت‌های داده‌کاوی و آنالیز آماری

داده‌کاوی	آنالیز آماری
-----------	--------------

<p>از فرضیه استفاده نمی‌شود.</p> <p>روابط در داده‌کاوی توسط الگوریتم‌های داده‌کاوی ایجاد می‌شود.</p> <p>داده‌ها عددی و غیر عددی (متنی) هستند.</p> <p>داده‌های نابجا و نادرست غیر قابل تشخیص هستند و باید داده‌ها درست باشند.</p>	<p>آمارشناسان از فرضیه‌ها استفاده می‌کنند.</p> <p>رابطه‌های آمارشناسان به فرضیه‌های آنان مربوط است.</p> <p>داده‌ها عددی هستند.</p> <p>تشخیص داده‌های نابجا و نادرست انجام می‌شود.</p>
--	---

سه ویژگی مهم داده‌کاوی عبارت است از:

- ذخیره و مدیریت تمام داده‌های موجود در بانک اطلاعات.
- امکان دسترسی تحلیلی گران و متخصصان به داده‌ها.
- تجزیه و تحلیل داده‌ها با استفاده از ابزارهای کامپیوتری.

نتیجه‌گیری

۸-۱-

سامانه‌های تصمیم‌یار، سیستم‌های هوشمند مبتنی بر اطلاعات هستند که با بکارگیری الگوها و روش‌های هوشمند تصمیم‌گیری و نیز مدل‌ها، داده‌ها، اطلاعات و دانش مستخرج از داده‌های مختلف، جهت یاری رساندن به مدیران جهت حل مسائل مختلف کمک می‌کنند. هیچ‌گاه یک سیستم تصمیم‌یار بجای مدیر تصمیم نمی‌گیرد بلکه به او در امر تصمیم‌گیری یاری می‌رساند. برای انجام هر پروژه، باید ایده خاصی را تعریف کنیم. جهت تعریف ایده خاص نیز باید از الگوهای مختلف فرایندها کمک گرفت تا نتیجه مطلوبی پدیدار شود.

۹-۱- ساختار پایان‌نامه

این پایان‌نامه در پنج فصل تهیه شده است:

فصل اول، به توضیح صورت مسئله و موضوعات کلی پایان‌نامه می‌پردازد. سیستم هوشمندی را تشریح می‌کند که قادر به تصمیم‌گیری است و ساختار این نرم‌افزار را تشریح و ابزارهای مورد نیاز جهت پیاده‌سازی آن را معرفی می‌کند. در این فصل اهداف و ساختار کلی پژوهش بررسی می‌گردد.

فصل دوم، بر مفاهیم اساسی و اولیه مورد نیاز تمرکز می‌کند و در حالت کلی به توصیف اجمالی مسئله پرداخته و روش‌ها و ابزارهای پیشنهادی را عنوان می‌کند. بهترین ابزار و روش را انتخاب و توضیحات کلی بیان می‌شود. الگوریتم‌ها و روش‌های آماری که مورد استفاده قرار گرفته-