

جستجوی اطلاعات مبتنی بر مدل های اطلاعاتی WP- AWP- AWPS بوسیله سیستم فازی

حسن توکلی

htzz69118@yahoo.com

دانشگاه آزاد اسلامی واحد کرج

زینب زارعی

htzz69118@yahoo . com

دانشگاه آزاد اسلامی واحد کرج

چکیده :

هدف از این مقاله چگونگی اعلام موجودیت اطلاعات مورد نیاز کاربر بر اساس خواسته های متنی مورد نظر می باشد . برای این منظور مدل های اطلاعاتی WP- AWP- AWPS معرفی می شوند و ضمن بررسی ساختار گسترده عملهای جستجوی اطلاعات در یک فضای اطلاعاتی ، چگونگی اعلام موجودیت اطلاعات خواسته شده بررسی می گردد و در نهایت چگونگی این اعلام بر اساس موتور فازی انجام می گردد .

واژگان کلیدی :

سیستم آگاه ساز ، مدل اطلاعاتی WP ، مدل اطلاعاتی AWP ، مدل اطلاعاتی AWPS ، سیستم های چند عاملی

۱- مقدمه :

گسترش روز افزون اطلاعات منجر به مدیریت اطلاعات در مراحل گردآوری ، تدوین و بازیابی اطلاعات می شود . با توجه به اینکه مرتب سازی اطلاعات در مرحله گردآوری خیلی مهم است ، نحوه تدوین و بازیابی اطلاعات که در این مقاله بطور خلاصه به آن پرداخته می شود ، قسمت مهم و چالش انگیز کار مدیریت اطلاعات می باشد . در این پروژه یک دایره بزرگ اطلاعاتی مانند یک کتابخانه دیجیتال مورد بررسی قرار داده می شود و در مورد تدوین و بازیابی اطلاعات موجود در این کتابخانه فرضی بحث می شود . [۱] در این مقاله منظور از اطلاعات ، اطلاعات متنی^۱ می باشد که به عنوان نمونه بررسی می گردد .

در ادامه مدل اطلاعاتی WP^۲ که به صورت متن آزاد^۳ و مبتنی بر اپراتورهای فاصله می باشد ، مورد بررسی قرار می گیرد. بعد از معرفی مدل اطلاعاتی WP به مدل اطلاعاتی AWP^۴ می رسیم . مدل اطلاعاتی AWP که در واقع مدل تکمیل شده ای از WP می باشد اساساً بر پایه مفهوم هم معنی بودن کلمات مورد بررسی واقع می شود. و نهایتاً به معرفی مدل اطلاعاتی AWPS^۵ می رسیم که مبتنی بر هم خانوادگی و هم معنی بودن کلمات مورد جستجو می باشد . [۲] در انتهای مقاله با استفاده از یک موتور جستجوی فازی و استفاده از مدل اطلاعاتی تکمیل شده AWPS ، نحوه بازیابی اطلاعات را تشریح می کنیم .

1- Textual 2- Word Pattern 3- Free text
4- Attribute Word Pattern 5- Attribute Word Pattern Similarity

۲- سیستم آگاه ساز با اطلاعات توزیع شده (Distributed Information Alert System : DIAS)

در یک سیستم آگاه ساز در جواب یک درخواست یک بسته اطلاعاتی به درخواست کننده فرستاده می شود ، در یک دایره گسترده بین کاربرها و منابعی که اطلاعات از آن بازیابی می شود ، جهت بهبود وضعیت ترافیک انتقال اطلاعات می توان از ایده عاملها (Agents) و به تناسب آن از عامل های هوشمند (Intelligent Agents) استفاده نمود .

استفاده از عاملهای هوشمند (Intelligent Agents) از جهت دیگری نیز مناسب است از آن جمله که می توان از عاملهای هوشمند میانی (Middle Intelligent Agents) جهت ساختن کاربر از رسیدن پیامهای منتشر شده به مقاصد دلخواه بهره جست . [۳و۴و۵]

ویژگی های سیستم آگاه ساز عبارتند از :

- سرویس دهی گسترده به معنی توانایی در برطرف ساختن نیازهای همه جانبه کاربر (از لحاظ جغرافیایی - موضوعی - ...)
- قواعد به معنی نگاشتی از حوزه حروف به حوزه اعداد توسط مدلهای اطلاعاتی WP AWP- AWPS
- تطابق به معنی تطبیق درخواست با اطلاعات بازیابی شده و بیان دو موضوع مهم ؛ کارآمدی و رضایت بخشی کاربر .
- فیلتر کردن جهت تطبیق بیشتر بین درخواست و اطلاعات بازیابی شده .
- طبقه بندی نمودن یک سری بسته های اطلاعاتی بازیابی شده با توجه به درخواست کاربر . [۲و۶]

1-2 الگوی کلماتی (Word Pattern : WP)

اساس WP مبتنی بر خواص منحصر به فرد یک الگوی کلماتی مورد جستجو و بیان این خاصیت منحصر به فرد توسط اپراتورهای فاصله و مجموعه قواعد مدل اطلاعاتی WP می باشد . به طور کلی الفبای مدل های WP AWP- AWPS مجموعه ای از فاصله های I می باشد که به صورت زیر مشخص می گردد .

$$I = \left\{ [1, u] : 1, u \in \mathbb{N}, 1 \leq u \right\} \cup \left\{ [1, \infty) : 1, u \in \mathbb{N} \text{ and } 1 \geq 0 \right\} \quad (1)$$

در واقع مجموعه I تعریف شده در فرمول (۱) دارای دو زوج مرتب I و u می باشد که I بیانگر کمترین فاصله دو الگوی کلماتی و u نیز بیانگر بیشترین فاصله دو الگوی کلماتی است . به عنوان مثال :

$$WP_1 <_{[1, u]} WP_2 \quad (2)$$

و بدین معنی می باشد که کلمه WP₂ در کمترین فاصله I و بیشترین فاصله u از WP₁ قرار دارد . در نهایت واژگان کلیدی یک متن به صورت زیر به هم متصل می گردند .

$$WP_1 <_{i_1} WP_2 <_{i_2} \dots <_{i_{n-1}} WP \quad (3)$$

که i ها زیر مجموعه ، مجموعه I می باشد که در (۱) معرفی شده اند . به عنوان مثال به نمونه های زیر می پردازیم :

کلمه کلیدی Neural Network را می توان به صورت

$$\text{Neural} <_{\{0, 0\}} \text{Network}$$

یعنی اینکه کلمه Neural همیشه بعد از کلمه Neural استفاده می شود . و یا دو کلمه Fuzzy Neural logic و Fuzzy logic را می توان به صورت :

$$\text{Fuzzy} <_{\{0, 6\}} \text{logic}$$

بکار برد یعنی اینکه کلمه logic در یک جا بلافاصله بعد از Fuzzy آمده و در جای دیگر به فاصله ۶ کلمه از Fuzzy بیان شده است . [۷]

2-2 تناسب الگوی کلماتی (Attribute Word Pattern: AWP)

اساس AWP مبتنی بر WP و تناسب (هم معنی بودن) کلمات مورد جستجو و تعداد محدودی از اختصاراتی به نام اختصارات جهانی می باشد . این اختصارات گسترده بوده که در واقع قوانین زبان AWP می باشد . [۸]
به عنوان نمونه بیان جمله های زیر را می توان مورد استفاده قرار داد : " من به نوشته هایی از محمد محمدی علاقه مند . " که می توان به صورت زیر بیان نمود :

AUTHOR] (Mohammad $\langle_{[0,0]}$ Mohammadi)

و جمله : " من به نوشته هایی از محمدعلی محمدی و محمد محمدی علاقه مند . " که می توان به صورت زیر بیان نمود:

AUTHOR] (Mohammad $\langle_{[0,3]}$ Mohammadi)

3-2 تناسب الگوی کلماتی به وسیله تشابه کلمات (Attribute Word Pattern Similarity : AWPS)

اساس AWPS مبتنی بر AWP و تشابه (هم خانواده بودن) کلمات مورد جستجو و تعداد محدودی عملگر اضافه شده به AWP بوده و بیان تشابه ارزش دو متن S_q و S_d می باشد. و به صورت کسینوس زاویه ای که به وسیله بردارهای متقابل به هم به وجود می آیند بیان می شود .

$$\text{Sim} (S_q , S_d) = \text{Cos} (S_q , S_d) = \frac{S_q \cdot S_d}{\|S_q\| \cdot \|S_d\|} = \frac{\sum_{i=1}^N W_{qi} \cdot W_{di}}{\sqrt{\sum_{i=1}^N W^2_{qi}} \sqrt{\sum_{i=1}^N W^2_{di}}} \quad (4)$$

که در این فرمول W_{di} و W_{qi} وزن کلمه W_i در متن های به ترتیب S_q , S_d می باشد . طریقه های مختلفی برای ساختن این وزنها می توان پیشنهاد نمود . ساده ترین بیان این وزنها توسط بیان مفهوم ساده تکرار یک کلمه در یک متن می باشد و می توان W_{qi} را به صورت تکرار کلمه W_i در متن S_q بیان نمود .

به عنوان مثال فرض می کنیم که دو متن A و B موجود است ، کلمات مورد جستجو در این دو متن عبارتند از a و b . تعداد تکرار کلمه a در متن A را به صورت $W_{Aa} = 3$ و تعداد تکرار کلمه b در متن A را بصورت $W_{Ab} = 4$ و تعداد تکرار کلمه a در متن B را بصورت $W_{Ba} = 0$ و تعداد تکرار کلمه b در متن B را بصورت $W_{Bb} = 4$ بیان می کنیم .

با توجه به (4) $\text{Cos} (S_q , S_d) = 0.8$ حاصل می شود ، که این مقدار می تواند بیانگر دوری و یا نزدیکی دو متن S_d , S_q نسبت به هم باشد ، و همچنین می توان با استفاده از یک متن مجازی (غیر واقعی) برای یافتن یک متن واقعی اقدام نمود .

۳- منطق فازی

منطق فازی شبیه یک منطق چند ارزشی است . در این منطق به جای درست یا نادرست ، سیاه یا سفید ، صفر یا یک ، سایه های نامحدودی از خاکستری بین سیاه و سفید وجود دارد . تمایز عمده منطق فازی با منطق چند ارزشی در آنست که در منطق فازی ، حقیقت و حتی ذات مطلب می تواند نادقیق باشد . در منطق فازی ، مجاز به بیان جملاتی از قبیل " کاملاً درست است " یا " کم و بیش درست است " هستیم و حتی می توان از احتمال نادقیق مثل " تقریباً غیر ممکن " ، " نه چندان " و " به ندرت " نیز استفاده کرد . در اینجا نیز می توان از منطق فازی برای بیان جملات " این دو متن شبیه یکدیگر هستند . " و یا " این دو متن کم و بیش شبیه یکدیگر هستند . " و یا " این دو متن تفاوت کمی نسبت به هم دارند . " استفاده نمود .

1-3 - جستجوگر فازی :

در ادامه از این خاصیت AWPS که به صورت کسینوس زاویه ای که به وسیله بردارهای متقابل به هم به وجود می آید ، می توان در یک جستجوگر فازی استفاده نمود . به این نحو که با تعریف تابع عضویت مناسب ، این تابع را برابر کسینوس برداری تعریف شده S_q , S_d قرار می دهیم :

$$\mu_A = \text{Cos} (S_q , S_d) \quad (5)$$

2-3 - پایگاه داده :

پایگاه داده زیر را می توان به عنوان نمونه معرفی کرد .

1. If (input1 is Smal) then (output1 is Bad) (1)
2. If (input1 is Big) then (output1 is Good) (1)

در ادامه تعاریف مربوط به توابع عضویت (Smal , Big) در شبیه سازی آورده شده اند .

3-3- ورودی به جستجوگر فازی :

که ورودی (input) همان کسینوس برداری تعریف شده S_q, S_d است . یعنی :

$$\text{Input} = \text{Cos} (S_q , S_d) \quad (6)$$

4-3 - نتایج شبیه سازی :

برای بیان برتری های الگوهای کلماتی بیان شده در این مقاله در دایره اطلاعات توزیع شده ، توسط یک سیستم آگاه ساز شبیه سازی هایی انجام شده است .

در شکل ۱ ، ۲ و ۳ نتایج شبیه سازی مقایسه دو متن را بر حسب کسینوس برداری بینشان که به ترتیب مقدار کم ، متوسط و زیاد است بررسی نموده و یک مقداری فازی به خروجی نسبت داده شده .

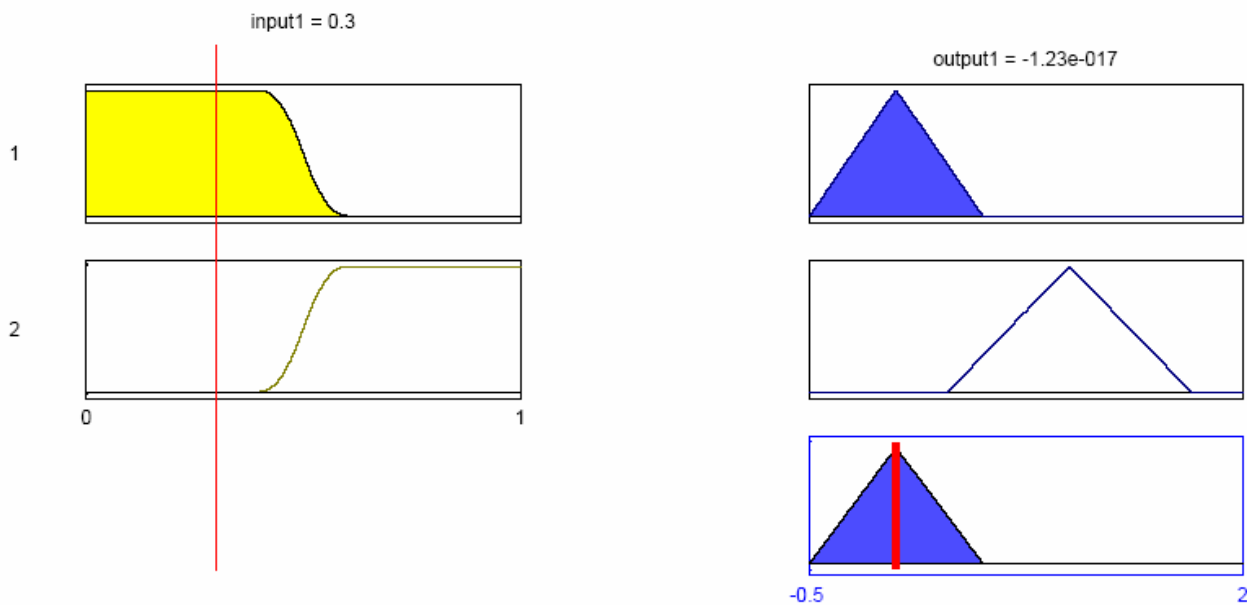


fig 1

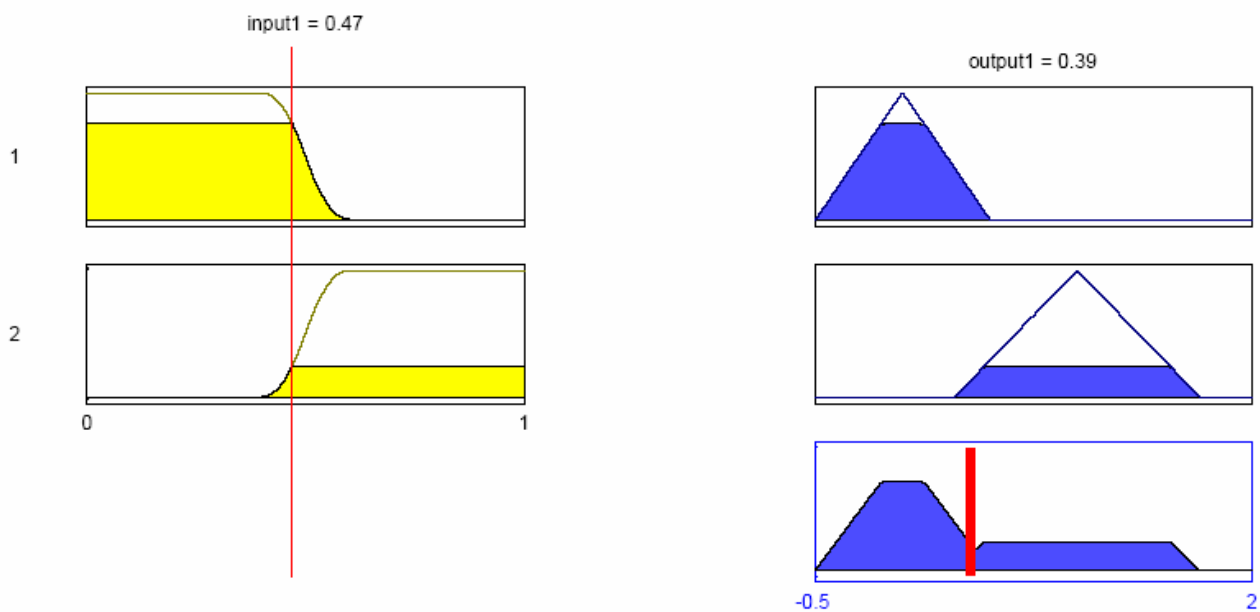


fig 2

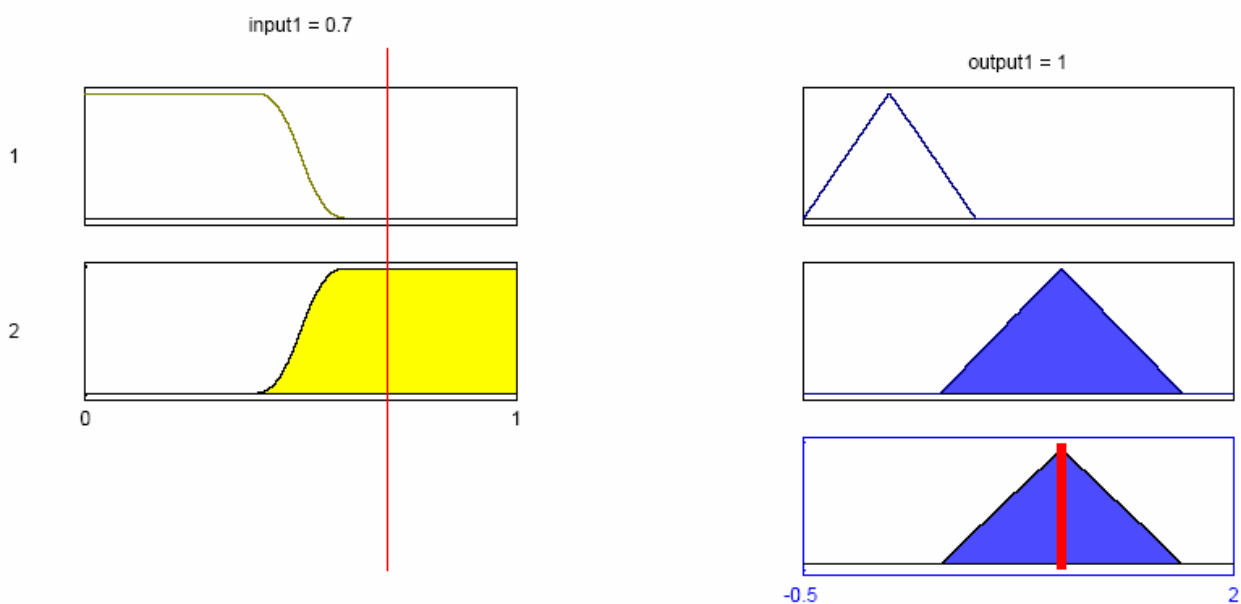


fig 3

۴- نتایج

در این مقاله به بررسی الگوهای جستجوی متنی در دایره اطلاعات توزیع شده توسط یک سیستم آگاه ساز و کاربرد آنها در سیستم های جستجو پرداخته شد. در انتها روشهای جستجوی فازی نوینی برای آن ارائه گردید. شبیه سازی های انجام شده کارایی بهتر این روش را نسبت به روشهای قبلی نشان می دهد.

منابع :

- [1] D.Faenses , L. Faulstich , H. Schweppe, A. Hinze , and A. Steidinger . Hermes – A Notification Service for Digital Libraries . In proceedings of the Joint ACM /IEEE Conference on Digital Libraries (JCDDL'01) , Roanoke , Virginia , USA ,2001
- [2] A.Carzaniga . Architectures for an Event Notification Service Scalable to Wide – area Networks . PhD thesis , Politecnico di Milano Italy, 1998 .
- [3] T. Finin , R.Fritzson, D.McKay , and R. McEntire. KQML as an Agent Communication Language . In N. Adam , B. Bhargava , and Y. Yesha , editors , Proceedings of the 3rd International Conference on Information and Knowledge Management (CIKM'94) , pages 456-463 , Gaithersburg , MD, USA,1994 .ACM Press .
- [4] P. Raftopoulou M. Koubarakis , C. Tryfonopoulos and T. Koutris . Data models and languages for agent – based textual information dissemination . In Proceedings of the 6th International Workshop on Cooperative Information Agents (CIA2002) ,Madrid , Lecture Notes in Computer Science . Springer , 2002 , forthcoming .
- [5] F. Wang . self – organising Communities Formed by Middle Agents . In Proceedings of the 1st International Joint Conference on Autonomous Agents & Multiagent Systems (AAMAS 2002) ,September 15-19 2002.
- [6] C.-C. K. Chang, H. Garcia –Molina, and A. Paepcke. Boolean Query Mapping across Heterogeneous Information Sources . IEEE Transactions on Knowledge and Data Engineering , 8(4):515-521, 1996.
- [7] C.-C. K. Chang, H. Garcia –Molina, and A. Paepcke. Predicate Rewriting for Translating Boolean Queries in a Heterogeneous Information System . ACM Transactions on Information System, 17(1): 1-39, 1999
- [8] M.F. Porter . An Algorithm for Suffix Striping . Program , 14(3) : 130-137,1980 .