

تسهيل دسترسي و استفاده از منابع زيست اطلاعاتي

نويسنده: جان. سي. بارتلت [۱]

ترجمه: مريم اسدي [۲] و اکرم اسدي [۳]

مقدمه

همانطور که طرح ژنوم انساني [۴] رو به اتمام است، گستره وسيعي از منابع زيست اطلاعاتي در حال گسترش يافتن است. اين مجموعه متشکل از پايگاههاي اطلاعاتي، از قبيل اطلاعات زيست شناختي و نرم افزارهاي تحليل و آماده سازي داده ها مي باشند.

جالب اينکه منابع زيست اطلاعاتي در هدايت پژوهشهاي زيست پزشيكي بسيار مؤثر بوده است. با اين همه، پژوهشهاي بسيار کمي در خصوص چگونگي استفاده متخصصان از اين منابع صورت گرفته شده است. اين مقاله طرح پيشنهاد شده براي پژوهش پايان نامه اي، به منظور دست يافتن به يك رويکرد، در تسهيل دسترسي به منابع زيست اطلاعاتي را ارائه مي کند.

زمينه

داده هاي توليد شده از بررسيهاي زيست شناسي مولکولي به توسعه دامنه وسيعي از پايگاههاي اطلاعاتي زيست شناسي مولکولي، پايگاههاي اطلاعاتي غيرکتابشناختي از اطلاعاتي مانند توالي DNA (براي مثال GenBank)، توالي پروتئين (براي مثال SwissProt) و اطلاعات نقشه ژنوم (براي مثال Genome Database) منجر شده است. توسعه، مديريت و استفاده از اين پايگاههاي اطلاعاتي يکي از مؤلفه هاي حوزه زيست اطلاعات و رشته مديريت داده ها به کمک رايانه است که به ما در گردآوري، تحليل و ارائه اين اطلاعات کمک مي کند (Persidis, 1999:828).

طرح ژنوم انساني در ۱۹۹۰، با هدف تعيين توالي تمام ژنوم انساني، تعيين محل ۱۰۰۰۰۰ ژن تخمين زده شده و ايجاد نقشه ژنتيکي و طبيعي از ژنوم انساني، آغاز گرديد (National Human Genome Research Institute, 2000). اتمام يك "پيش نويس کاري" از توالي ژنوم انساني در ۲۶ جون سال ۲۰۰۰ اعلام شد و پيش بيني شد يك سوم ژنوم بطور کامل تا سال ۲۰۰۱ تعيين توالي شوند. هدف فعلي، تعيين کامل توالي ژنوم انساني تا سال ۲۰۰۲ است، دو سال قبل از وقت مقرر که در سال ۱۹۹۰ در نظر گرفته شده است (Baxevanis, 2000:1).

آخرين شماره سالنامه پايگاه اطلاعاتي از مجله Nucleic Acids Research بيش از ۲۰۰ پايگاه اطلاعاتي زيست شناسي مولکولي را فهرست نمود (Baxevanis, 2000). اطلاعات در اين پايگاهها متشکل از توالي هاي DNA، RNA و پروتئين، ساختمان ژن، نقشه ها، جهش ها و ژنوم ها از جمله اين موارد مي باشند. به تنهائي پايگاه اطلاعاتي GenBank به طور متوسط در روز، دو ميليون جفت پايگاه افزايش مي يابد (Baxevanis, 2000). نظام ENTREZ کتابخانه ملي پزشيكي، جستجوي اصلي واسط کاربري خود را هم به پايگاه اطلاعاتي کتابشناختي (براي مثال Medline, HealthSTAR) و هم به پايگاه اطلاعاتي زيست شناسي مولکولي (براي مثال GenBank)، که ۵۰۰۰۰ بار در روز جستجو مي شود، مجهز کرده است (Persidis, 1999).

همانطور که تعداد و حجم پايگاههاي اطلاعاتي زيست شناسي مولکولي گسترش مي يابد، استفاده از آنها نيز افزايش مي يابد. در مقابل نگرانيهايي در خصوص کمبود مهارت در حوزه زيست اطلاعات رو به افزايش مي باشد (Collins et al., 1998; MacLean & Miles, 1999). متخصصان شديدا تحت تأثير گستردگي و حجم زياد منابع زيست اطلاعاتي قابل دسترس، براي آنها و غالبا کمبود مهارت براي استفاده از اين منابع قرار گرفته اند (Yarfitz & Ketchell, 2000). شواهد داستانه گونه نشان مي دهد، براي متخصصان غيرعادي نيست، که زمان طولاني در آزمایشگاه تحقيقاتي براي رسيدن به نتايج صرف نمايند، بلکه مي توانند با صرف زمان کم تري از منابع زيست اطلاعات جهت رسيدن به نتايج خود استفاده نمايند. استفاده از اين منابع مي تواند به صرفه جويي هم در وقت و هم در هزينه براي محققان منجر شود. پژوهش کمي در خصوص استفاده از اين منابع مفيد پژوهشي صورت گرفته است. با وجود گستردگي، پيچيدگي و حجم زياد و رو به گسترش منابع زيست اطلاعاتي، پژوهش در اين زمينه يعني براي چه نوع اطلاعاتي کاربران به اين منابع مراجعه مي کنند، چگونه اطلاعات خود را جستجو مي کنند و چگونه از اين اطلاعات

استفاده می کنند، بسیار مفید می باشد. نتایج تحقیقات در این حوزه ها می تواند برای اهداف زیادی استفاده شود. از آنها می توان در طراحی و توسعه منابع زیست اطلاعاتی به منظور پاسخ گویی به نیازهای کاربران بهره برد. همچنین این نوع پژوهش ها می تواند به عنوان پایه ای در گسترش برنامه های آموزشی و تعلیم و تربیت استفاده شود. به همین دلیل متخصصان ممکن است در استفاده از منابع زیست اطلاعاتی بهتر و بیشتر آگاهی و مهارت یابند. همچنین این پژوهش ها می تواند به توسعه ابزارهایی برای راهنمایی و کمک به متخصصان در زمانیکه آنها در حال استفاده از این منابع هستند، بیانجامد.

پژوهش زیادی در زمینه رفتار اطلاعاتی متخصصان علوم زیستی، زیست شناسان مولکولی در استفاده از منابع سنتی و کتابشناختی صورت گرفته است (Dillon, Curtis, Weller, & Hurd, 1993, 1997; Bayer & Jahoda, 1981; Rolinson, Meadows, & Smith, Palmer, 1991a, 1991b; Garvey, Tomita, & Woolf, 1974; French, 1990; ۱۹۸۱; ۱۹۹۵; Skelton, 1973). اما پژوهش کمی در خصوص استفاده از منابع زیست اطلاعاتی غیرکتابشناختی صورت گرفته است. بررسی ها بر روی رفتار اطلاع یابی زیست شناسان مولکولی نشان داد که، همانند اوائل سال ۱۹۹۹، همانطورکه تصور می شود پایگاههای اطلاعاتی توالی مولکولی دارای نقش بیش از پیش مهمی هستند، بلکه نیاز به مهارت در استفاده از پایگاههای اطلاعاتی وجود دارد یا مهارت در استفاده از آن را باید پشت سر بگذارند (Grefsheim et al., 1991). در همان زمان ضرورت کمک در مدیریت توسعه مجموعه ای از اطلاعات قابل دسترس بوجود آمد (Grefsheim et al., 1991). تقریباً ۱۰ سال بعد، مطالعه ای نشان داد که ۷۰ درصد زیست شناسان مولکولی تحت بررسی، در حال استفاده از پایگاههای اطلاعاتی توالی مولکولی بطور هفتگی یا ماهانه بودند (Yarfitz & Ketchell, 2000). در حالیکه، ارتباطات شخصی با همکاران، منبع اولیه اطلاعاتی در میان این منابع بود. همچنین علاقه قابل توجهی به خدمات مشاوره ای زیست اطلاعاتی، کلاسها و خدمات دیگر به منظور فراهم آوردن کمک در استفاده از منابع زیست اطلاعات مشاهده گردید (Yarfitz & Ketchell, 2000).

اهداف پژوهش

هدف پژوهش در دو بخش طراحی گردید: الف) ارزیابی الگوهای استفاده از منابع زیست اطلاعاتی (هم پایگاههای اطلاعاتی زیست شناسی مولکولی و هم ابزارهای نرم افزاری تحلیلی مرتبط) توسط کاربران خبره ب) ایجاد و آزمون یک ابزار بر پایه یک الگوی پیش‌بینانه، که دسترسی و استفاده از منابع زیست اطلاعاتی را توسط متخصصان تسهیل نماید. در این میان چالش های در پیش روی متخصصان از لحاظ انتخاب و استفاده مناسب از منابع زیست اطلاعاتی وجود دارد. به منظور گرفتن الگوهای استفاده، این پژوهش تحلیل خواهد کرد، چگونه خبرگان از منابع زیست اطلاعاتی استفاده می کنند. یعنی برای متخصصانی که در این پژوهش مشارکت دارند، آیا الگوها یا ویژگی های عمومی برای استفاده آنها از منابع زیست اطلاعاتی وجود دارد؟ نویسنده از این الگوها، الگویی پیش‌بینانه برای استفاده از منابع زیست اطلاعاتی برای یک وظیفه علمی خاص و استفاده از الگویی برای توسعه ابزاری به منظور کمک به متخصصان در انتخاب و استفاده از آنها فراهم خواهد نمود. دو هدف پژوهشی ویژه وجود دارد:

- ۱- شناسایی و ارزیابی الگوهای استفاده از منابع زیست اطلاعاتی توسط متخصصان، به دلیل آنکه یک الگوی پیش‌بینانه بتواند توسعه یابد.
- ۲- ایجاد و آزمون ابزاری مبتنی بر رایانه، بر مبنای الگوی پیش‌بینانه، که بتواند مبتدیان را در دسترسی و استفاده از منابع زیست اطلاعاتی کمک نماید.

طراحی پژوهش

مرحله اول

هدف: شناسایی و ارزیابی الگوهای استفاده از منابع زیست اطلاعاتی هدف اولین مرحله، شناسایی و ارزیابی الگوهای استفاده از منابع زیست اطلاعاتی توسط کاربران خبره، متخصصانی که در استفاده از منابع زیست اطلاعاتی دارای مهارت هستند، می باشد. متخصصان در تمام انواع پژوهش مشابه درگیر خواهند بود، زیرا که در وظایفی که نیاز به اطلاعات دارند، مشابه خواهند بود. بیست متخصص با استفاده از یک مصاحبه نیمه ساختاری، مورد پرسش قرار خواهند گرفت. متخصصان فرصت خواهند یافت که به شرح تجربیات خود با بیان خود و بحث درباره آنچه را که برای آنها اهمیت دارد، بپردازند. بطوریکه این فرصت به پژوهشگر این اطمینان را خواهد داد که نکات کلیدی مورد توجه قرار می گیرد. تمرکز در مصاحبه بلادرنگ به درک وظیفه علمی در استفاده از منابع زیست اطلاعاتی و فرآیندی که توسط آنها استفاده می شود، خواهد انجامید. در پرسش مصاحبه اولیه از متخصصان سؤال خواهد شد که به توصیف مراحل که در ذیل خواهد آمد، بپردازند، و پایگاههای اطلاعاتی و ابزارهایی که آنها به منظور موفقیت کامل در انجام یک وظیفه علمی استفاده خواهند نمود،

- بیان نمایند. پرسشهای بررسی به منظور اطمینان از اینکه نکات ضروری پوشش داده می شوند، استفاده خواهد شد. این پرسشها شامل موارد زیر خواهد بود:
- چه توالی از رویدادها از پی هم دنبال می شوند؟
 - چه پایگاههای اطلاعاتی و ابزارهایی استفاده می شوند؟
 - آغازگر این فرآیند چیست؟
 - در آغاز فرآیند دارای چه اطلاعاتی هستید؟
 - آیا اطلاعات بایستی به هر طریقی آماده سازی شوند؟
 - در پایان فرآیند چه اطلاعاتی منتشر می شود؟
 - بعد از این چه اتفاقی می افتد؟
 - چطور شما در می یابید چه زمانی فرآیند تمام می شود؟

پرسشهای بیشتر، تعدادی از متغیرهایی که در فرآیند توسط متخصصان توصیف شده، به منظور درک اینکه آیا فرآیند می تواند به سایر موقعیت ها تعمیم داده شود را بیان خواهد نمود. همچنین درک نقاط تصمیم گیری در فرآیند، چرا منابع خاص انتخاب می شوند و هدف از هر نوع تجزیه و تحلیل چیست، مهم است.

داده های گردآوری شده از مصاحبه ها به منظور گرفتن الگوهای در استفاده از منابع زیست اطلاعاتی تحلیل خواهد شد، یعنی در یک وظیفه علمی مشخص، منابع زیست اطلاعاتی دسترس پذیر می شوند و در یک الگوی یکپارچه مورد استفاده قرار می گیرند. از این الگوها، یک الگوی پیش‌بینانه توسعه خواهد یافت. این فرآیند تعاملی، الگویی در حال توسعه خواهد بود، به طوریکه داده ها تحلیل و بررسی می شوند و با هر دسته، داده های جدید تعریف می شوند. علاوه برای درک فرآیندهای دنبال شده توسط متخصصان، همچنین درک منابع متنوع بکار برده شده توسط آنان نیز ضروری خواهد بود. تحلیل منابع مستلزم یک رده بندی آرایه شناختی خواهد بود. درک هدف و کارکردهای پایگاههای اطلاعاتی متنوع و ابزارهای تحلیلی، در چارچوبی که چگونه هر کدام از پایگاههای اطلاعاتی استفاده می شوند، قابلیت های هر یک از آنها، تشابهات و تفاوتها، آنان مهم می باشد. خصوصا درک هدف و کارکردهای آنها برای درک اینکه منابع دارای شباهت هستند یا دارای کارکردهای همپوشانی و در چه چیزی تفاوت دارند، مهم خواهد بود. همچنین داده های مصاحبه به منظور درک دلایل بعد از فرآیندهای توصیف شده، تحلیل خواهد شد. این مورد نه فقط به منظور درک اینکه چه منابعی استفاده می شود، بلکه برای درک چگونگی و چرایی استفاده، مهم خواهد بود. تحلیل کیفی، تمرکز کمی را که منجر به توسعه الگو خواهد شد، افزایش خواهد داد.

مرحله دوم

هدف: توسعه و آزمون یک ابزار به منظور تسهیل دسترسی و استفاده از منابع زیست اطلاعاتی

دومین مرحله مستلزم ایجاد و ارزیابی ابزاری برای کمک به متخصصان به منظور دستیابی و استفاده از منابع زیست اطلاعاتی خواهد بود. ابزار بر پایه، الگوی مرحله اول توسعه خواهد یافت و می تواند چندین شکل در برداشته باشد. در این میان، یک راهنمای جستجو، ابزار حمایت از تصمیم گیری یا مواد آموزشی وجود دارد. یافته های حاصل از مصاحبه ها، به تعیین مناسبترین شکل ابزاری کمک خواهد نمود. شکل دقیق ابزار یا چیزی که آن را پوشش خواهد داد، شناخته نخواهد شد، مگر آنکه داده ها از اولین مرحله تحلیل شوند.

یکبار ابزار توسعه یافته با یک گروه متخصصانی که در علم زیست اطلاعات خبره نیستند، آزمون خواهد شد. به منظور تعیین اینکه، آیا آن ابزار می تواند دسترسی و استفاده از منابع زیست اطلاعاتی را تسهیل نماید. آزمون، الگویی را که براساس ابزار پایه گذاری میشود و هم خود ابزار را ارزیابی خواهد کرد، انجام می گیرد. با این وجود، تأکید بر روی آزمون و تعریف ارزش پیش‌بینانه الگو و کاربردش، بعنوان پایه ای برای یک ابزار خواهد بود، بجای اینکه براساس قابلیت استفاده ابزار باشد. برای شرکت کنندگان در آزمون یک وظیفه علمی استاندارد شده ای تعیین خواهد شد (مانند تولید یک نقشه محدود یا تشریح و توصیف توالی جدید). بطوریکه منابع زیست اطلاعاتی بتوانند با یک راه حل ارائه شوند. این مورد به حل مسئله منجر خواهد شد و بطور تصادفی یا استفاده از ابزار را تعیین خواهد کرد یا تعیین نخواهد کرد. در این میان عوامل مورد بررسی در ارزیابی کارآیی، اثربخشی و رضایت کاربر هستند.

نتایج

این پژوهش تعیین خواهد کرد، آیا کاربرد الگوها توسط متخصصان زیست اطلاعاتی به اندازه کافی به منظور کمک به دیگر متخصصان (متخصص غیرزیست اطلاعاتی) در دسترسی و استفاده از منابع زیست اطلاعاتی، پیش‌بینانه می باشند. پیش بینی می شود الگویی بتواند در موفقیت انجام یک وظیفه علمی دنبال شوند، وجود خواهد داشت. یک مثال ساده از یک وظیفه، تعیین و چگونگی حذف یک قطعه خاص از DNA کروموزوم یک باکتری است. ابتدا بایستی تمامی توالی کروموزوم شناسایی شوند. این اطلاعات قابل دسترس در پایگاههای اطلاعاتی توالی یافت

می شوند. سپس توالی بازسازی شده بایستی بوسیله یک برنامه رایانه ای تحلیل شوند، بطوریکه جستجوها برای محل و نقاط خاص روی DNA که توسط آنزیم های تخصصی جدا خواهند شد، شناسایی می شوند. خروجی شامل تحلیل نقشه ای از تمام محل های بریده شده بر روی کروموزوم می باشد. سپس پژوهشگران این محلها را شناسایی خواهند کرد و از این طریق می توانند قطعه ای که قرار است حذف شود شناسایی نمایند.

این نمونه، به وضوح به توصیف یک وظیفه علمی ساده می پردازد، معهذاً نوعی از الگوهای را که ممکن است از نحوه کاربرد الگوهای توسط متخصصان آشکار شود را توصیف می کند. پایگاه اطلاعاتی و ابزار تحلیلی استفاده شده، ممکن است در میان متخصصان متنوع باشد، انتخاب ممکن است توسط چندین عامل شامل جزئیات خاص وظیفه (برای مثال، موجود زنده در حال مطالعه)، منابع قابل دسترس و استنتاج فردی مؤثر باشد. به هر حال رده منبع استفاده شده، قابل پیش بینی خواهد بود.

انتظار می رود این پژوهش به توسعه ابزاری برای کمک به متخصصان در کاربرد منابع زیست اطلاعاتی برای انجام وظیفه خاص منجر شود. این می تواند به یک راه حل برای مسئله متخصصان که تحت تأثیر دامنه ای از منابع قابل دسترس هستند فراهم کند و به آنها در ردیابی در میان آن منابع کمک نماید. همچنین این پژوهش امکان سنجی استفاده از الگوی پیش بینانه از توسعه از طریق بررسی کاربران خبره، ایجاد یک ابزار حمایتی تصمیم گیری، راهنمای جستجو یا مواد آموزشی را بررسی خواهد نمود. اگر اثربخش بودن بررسی درک شود، این رویکرد می تواند در جهت وظایف علمی دیگر که در آنها منابع زیست اطلاعاتی بکار برده می شوند، استفاده گردد. همچنین این رویکرد می تواند در رشته های دیگر با منابع اطلاعاتی مشابه استفاده شود.

یادداشتها:

[1] - Bartlett, Joan C. (2001) "Facilitating access to and use of bioinformatics resources". Information Research, 6(2) Available at: <http://InformationR.net/ir/6-2/ws1.html>

[۲] - کارشناس ارشد کتابداری و اطلاع رسانی

[۳] - کارشناس ارشد زیست شناسی

[4] - Human Genome Project

منابع

- Baxevis, A. D. (2000). "The Molecular Biology Database Collection: an online compilation of relevant database resources." *Nucleic Acids Research*, 28(1), 1-7.
- Bayer, A., & Jahoda, G. (1981). "Effects of online bibliographic searching on scientists' information style." *On-Line Review*, 5, 323-33.
- Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., & Walters, L. (1998). "New goals for the U.S. Human Genome Project: 1998-2003." *Science*, 282(5389), 682-9.
- Curtis, K. L., Weller, A. C., & Hurd, J. M. (1993). "Information-seeking behaviour: a survey of health sciences faculty use of indexes and databases." *Bulletin of the Medical Library Association*, 81(4), 383-392.
- Curtis, K. L., Weller, A. C., & Hurd, J. M. (1997). "Information-seeking behavior of health sciences faculty: the impact of new information technologies." *Bulletin of the Medical Library Association*, 85(4), 402-10.
- Dillon, M. (1981). "Serving the information needs of scientific research." *Special Libraries*, 72, 215-23.
- French, B. A. (1990). "User needs and library services in agricultural sciences." *Library Trends*, 38, 415-41.
- Garvey, W. D., Tomita, K., & Woolf, P. (1974). "The dynamic scientific information user." *Information Storage and Retrieval*, 10, 115-31.
- Grefsheim, S., Franklin, J., & Cunningham, D. (1991). "Biotechnology awareness study, part 1: where scientists get their information." *Bulletin of the Medical Library Association*, 79(1), 36-44.
- MacLean, M., & Miles, C. (1999). "Swift action needed to close the skills gap in bioinformatics." *Nature*, 401, 10.
- National Human Genome Research Institute. (2000). "Human Genome Project Goals: 1998-2003." Bethesda, MD: Retrieved March 24, 2000 from the World Wide Web: <http://www.nhgri.nih.gov/HGP/>.
- Palmer, J. (1991a). "Scientists and information: I. Using cluster analysis to identify information style."

Journal of Documentation, 47(2), 105-29.

- Palmer, J. (1991b). "Scientists and information: II. Personal factors in information behaviour." *Journal of Documentation*, 47(3), 254-75.
- Persidis, A. (1999). "Bioinformatics." *Nature Biotechnology*, 17, 828-30.
- Rolinson, J., Meadows, A. J., & Smith, H. (1995). "Use of information technology by biological researchers." *Journal of Information Science*, 21(2), 133-9.
- Skelton, B. (1973). "Scientists and social scientists as information users: a comparison of results of science user studies with the investigation into information requirements of the social sciences." *Journal of Librarianship*, 5(2), 138-56.
- Wadman, M. (1999). "Human Genome Project aims to finish 'working draft' next year." *Nature*, 398, 177.
- Yarfitz, S., & Ketchell, D. S. (2000). "A library-based bioinformatics services program." *Bulletin of the Medical Library Association*, 88(1), 36-48.