

## نظام نمایه‌سازی "وورداسمیت" در شبکه کتابشناختی (او.سی.ال.سی)

نوشته: فاطمه معتمدی\*

این شبکه، از کارآئی و توانائی کافی و لازم در بازیابی سریع و موثر اطلاعات از محیط شبکه نیز برخوردار باشد.

### چکیده

کلیدواژه‌ها: شبکه کتابشناختی (او.سی.ال.سی)، نمایه‌سازی (دبلیو.اس)، نمایه‌سازی خودکار، پردازش زبان طبیعی

### مقدمه

نمایه مهمترین ابزار در نظامهای بازیابی اطلاعات و کلید اساسی برای دستیابی به اطلاعات مندرج در منابع اطلاعاتی است. ارزش هر نمایه در گرو تهیه بهترین نقاط دستیابی و تأمین رضایت استفاده‌کنندگان در بازیابی اطلاعات است.

سالهاست که کتابداران مسائل مربوط به ذخیره و بازیابی اطلاعات را مدنظر قرار داده و برای کنترل موثر مواد سنتی، ابزاری را توسعه داده‌اند. توجه به روند تکاملی فرآیند نمایه‌سازی مبین این نکته است که افزایش حجم انتشارات و تغییر در شکل ارائه خدمات اطلاع‌رسانی، نظام‌های نمایه‌سازی را نیز دستخوش تحولات قابل توجهی نموده است. ابداع انواع نمایه‌های پیش‌همارا، پس همارا، نمایه‌های خودکار و استفاده از انواع زبان‌های طبیعی، آزاد و ساخت یافته یا کنترل شده همه تلاش‌هایی

رشد سریع اطلاعات الکترونیکی در عصر حاضر، زمینه‌ساز ایجاد انواع متعدد شبکه‌های کتابشناختی گردیده است. دستیابی به اطلاعات الکترونیکی مستلزم استفاده از روش موثر در نمایه‌سازی و تشخیص کلیدواژه‌های مهم بیانگر محتوای یک مدرک است. در همین راستا، (او.سی.ال.سی)<sup>۱</sup> که بزرگترین شبکه کتابشناختی جهان محسوب می‌شود به منظور فراهم آوردن تسهیلات لازم برای دستیابی سریع و موثر به اطلاعات، اقدام به ساخت سیستم‌نمایه‌سازی خودتحت عنوان "وورد اسمیت" نموده است. مبنای کار این سیستم، تشخیص خودکار مفاهیم، کلمات و عبارات کلیدی در متن ماشین‌خوان، براساس «پردازش زبان طبیعی» است. به این ترتیب شبکه (او.سی.ال.سی) با استفاده از تئوری‌های زبان‌شناسی محاسباتی و معنائی-که مورد تأکید پردازش زبان طبیعی است-و نیز با بکارگیری تئوری‌های اطلاع‌رسانی و مدل‌های آماری، سعی در ساخت نمایه‌ای داشته است که علاوه بر همخوانی با حجم عظیم داده‌های

\* عضو هیئت علمی دانشگاه علوم پزشکی کرمان و دانشجوی دکتری علوم کتابداری و اطلاع‌رسانی دانشگاه شیراز

بوده‌اند در جهت بهبود و تکمیل فرآیند نمایه‌سازی در انجام رسالت خاص خود.

ایجاد شبکه‌های متعدد کتابشناختی و نیز رشد سریع آن‌ها، فرصت‌های متعدد و چالش‌های بی‌سابقه‌ای را فراروی حرفه کتابداری و اطلاع‌رسانی برای دستیابی به حجم عظیمی از اطلاعات گذارده است. چنین شرایطی نیاز به مجتمع و یکپارچه‌سازی منابع موجود در محیط وب و اینترنت همراه با انواع مواد سنتی کتابخانه‌ای، ضرورت بررسی مجدد ابزارهای موجود در بازیابی اطلاعات و کنترل کتابشناختی را ایجاب می‌نماید.

بزرگترین شبکه کتابشناختی پیوسته موجود، شبکه (ا.وسی.ال.سی) می‌باشد (هسی-یی، ۱۹۹۶) که در سال ۱۹۶۷ ایجاد شده است و حاوی بیش از ۳۱ میلیون رکورد در فرمت‌های مختلف است. این شبکه یک سازمان تحقیقاتی غیرانتفاعی واقع در اوهایو می‌باشد که با ارائه خدمات کتابخانه‌ای امکان دستیابی ۴۱ هزار کتابخانه مختلف را به اطلاعات فراهم می‌آورد. هدف کلی (ا.وسی.ال.سی) افزایش دسترس‌پذیری منابع کتابخانه‌ای از طریق اشتراک در منابع و کاهش نرخ رشد هزینه‌های کتابخانه‌ای است. شبکه (ا.وسی.ال.سی) در راستای تحقق بخشیدن به این هدف اقدام به تهیه پروژه‌هائی نموده است که یکی از مهمترین این پروژه‌ها، پروژه نمایه‌سازی دلبیواس می‌باشد.

هدف نمایه‌سازی دلبیواس اصلاح و بهبود دستیابی به اطلاعات الکترونیکی با استفاده از روشهای مؤثر برای تشخیص کلیدواژه‌های مربوط به محتوای یک مدرک می‌باشد. تشخیص کلمات و عبارات باید به گونه‌ای باشد که به بهترین نحوی مبین موضوع مدرک باشند و در نتیجه نمایه‌های مفید و کارآمدی برای مدارک موجود در وب ایجاد گردد. به همین منظور مبنای این نمایه‌سازی، تشخیص خودکار مفاهیم در متن ماشین‌خوان براساس پردازش زبان طبیعی، مدل‌های آماری، زبانشناسی محاسباتی و تئوری رده‌بندی می‌باشد.

مقاله حاضر ضمن معرفی شبکه کتابشناختی (ا.وسی.ال.سی) و اشاره به خدمات، پایگاه‌های اطلاعاتی و پروژه‌های این شبکه، بطور اجمالی به مرور نظام‌های

نمایه‌سازی و زبان‌های بازیابی اطلاعات می‌پردازد و سپس طراحی و ساختار نظام نمایه‌سازی دلبیواس را تشریح می‌نماید.

### پایگاه کتابشناختی (ا.وسی.ال.سی)

انواع پایگاه‌های کتابشناختی با هدف تسهیل و تسریع در ارائه خدمات کتابشناختی و کمک به پیشرفت دانش ایجاد شده‌اند و از بدو پیدایش خود تاکنون با ارائه سرویس‌ها و خدمات متعدد، افراد و کتابخانه‌های مختلف را در تحقق اهداف خود یاری بخشیده‌اند. این پایگاه‌ها دامنه وسیعی از خدمات را ارائه می‌دهند از جمله خدمات تهیه مواد، فهرست‌نویسی و خدمات مرجع. بسیاری از کتابخانه‌ها به دلیل فشارهای اقتصادی، رشد روزافزون حجم انتشارات و در نتیجه لزوم وجود شبکه‌ای برای اشتراک منابع و تعاون بین کتابخانه‌ها و همچنین لزوم تغییر فاز از مالکیت به سمت دستیابی به اطلاعات، بر استفاده از این شبکه‌ها تأکید دارند. بطور کلی از طریق این پایگاه‌ها می‌توان به سه پرسش اساسی پاسخ داد: (۱) سئوالات کتابشناختی یعنی امکان بدست آوردن اطلاعات کامل درباره یک کتاب، مجله یا انواع خاص دیگر مواد کتابخانه‌ای. سیستم‌های شبکه‌ای معمولاً داده‌های فهرست‌نویسی کاملی بر روی صفحه (سی.آر.تی) پایانه می‌آورند. (۲) سئوالات تحقیقی (تعیین صحت و سقم چیزی) مربوط به تشخیص هجی صحیح نام یک نویسنده، نام ناشر، تاریخ و غیره. (۳) سؤال در مورد تعیین محل مدرک. از طریق این قبیل سیستم‌ها می‌توان فهمید که کدام کتابخانه‌ها مواد اطلاعاتی مورد نیاز را دارند.

سیستم‌های کتابشناختی مثل (آر.آل.آی.ان)، (ا.وسی.ال.سی) و غیره شبکه‌هائی غیرانتفاعی و بین‌المللی هستند که با روش‌های متنوعی اداره می‌شوند و به شکل‌های بسیار متفاوت اقدام به ارائه خدمات می‌نمایند. عملکردهای خاص این شبکه‌ها عبارتند از: تعیین محل سریع کتاب، مقاله و غیره؛ فراهم‌آوری و تهیه مواد از طریق مستقیم یا غیرمستقیم؛ کنترل گردش و دستیابی فهرست راه دور و فهرست‌نویسی مشترک. هر سیستم شامل مدخل‌های فهرستگان ملی مثلاً رکوردهای مارک

(جردان، ۱۹۹۸). این شورا متشکل از کتابداران و سایر شرکائی است که شایستگی و صلاحیت آن‌ها برای عضویت منوط به میزان سهمی است که در فهرستگان آنلاین (اوس.ال.سی) دارند. با ایجاد اصلاحات در سخت‌افزار، امکانات بیشتری برای استفاده‌کنندگان بیشتری فراهم نمود. از سال ۱۹۷۳ کتابخانه‌های غیرآکادمیک توانستند عضو (اوس.ال.سی) شوند و نیز بر تعداد شبکه‌های خارج از ایالتی که به (اوس.ال.سی) پیوستند افزوده گشت.

کتابخانه‌ها با عضویت در (اوس.ال.سی) می‌توانند به همه اطلاعات جهانی و به تمامی خدمات و پایگاه‌های اطلاعاتی دسترسی یابند. این عضویت، بزرگترین کنسرسیوم کتابخانه‌های جهانی را در بر می‌گیرد و در بین سایر شبکه‌های کتابشناختی بیشترین تعداد اعضا را دارد. تعداد اعضای آن تا سال ۱۹۹۱ بیش از ۷۰۰۰ کتابخانه در سراسر دنیا بود که اکثر آن‌ها از امریکا و کانادا بودند.

در سراسر تاریخ (اوس.ال.سی) ارتباطات راه دور، یک وسیله مهم برای تداوم اهداف کلی خود در زمینه افزایش دستیابی به اطلاعات جهانی و کاهش هزینه اطلاعات بوده است (جردان، ۱۹۹۸). این شبکه همچنان براساس پروتکل استاندارد صنعتی (تی.سی.پی.آی.پی) طی سه سال آینده در جهت تبدیل به یک شبکه کاملاً باز به پیش می‌رود. از طرف دیگر امکان دستیابی با کیفیت بالا را برای کتابخانه‌ها فراهم می‌سازد.

هدف کلی و اساسی (اوس.ال.سی) که بر مبنای افزایش سهولت دستیابی و استفاده از پیکره رو به رشد دانش و اطلاعات علمی، ادبی و آموزشی (گزارش سالیانه، ۱۹۹۵) قرار دارد، عبارت است از افزایش دسترس‌پذیری منابع کتابخانه‌ای و کاهش نرخ رشد هزینه‌های کتابخانه. در راستای تحقق این هدف، اهداف ذیل را دنبال می‌کند:

استقرار، نگهداری و پیاده کردن یک شبکه کتابخانه‌ای کامپیوتری و افزایش استفاده از کتابخانه، بررسی و توجه به مسائل مربوط به کتابخانه‌ها و کتابداری، فراهم نمودن امکانات و محصولات که به نفع استفاده‌کنندگان کتابخانه و کتابخانه‌ها باشد؛ از جمله افزایش قابلیت دسترسی به منابع کتابخانه برای مراجعین به کتابخانه‌ها و کاهش هزینه‌های هر واحد کتابخانه.

است. این رکوردها شامل همه فرمت‌ها از جمله کتاب، نشریه، ادواری و نسخ دستی می‌باشند.

بزرگترین شبکه کتابشناختی موجود، شبکه (اوس.ال.سی) می‌باشد (هسی، ۱۹۹۶) که به منظور دستیابی بیشتر به اطلاعات جهانی بوجود آمده است. مؤسسات و کتابخانه‌های دانشگاهی و تحقیقاتی (هر کدام با توجه به نوع نیاز خود) در سطح وسیعی از این شبکه استفاده می‌کنند. این شبکه، یک سازمان تحقیقاتی غیرانتفاعی است که به ارائه خدمات کتابخانه‌ای می‌پردازد و در اوهایو واقع شده است.

### تاریخچه، اهداف و عضویت در (اوس.ال.سی)

در سال ۱۹۶۷، مسئولین دانشکده‌ها و دانشگاه‌ها در اوهایو، یک سیستم کامپیوتری به نام «مرکز کتابخانه‌ای کالج اوهایو»<sup>۲</sup> را ایجاد نمودند که در آن کتابخانه‌های مؤسسات آکادمیک اوهایو می‌توانستند در منابع مشترک شوند و هزینه‌ها را کاهش دهند. اولین دفاتر (اوس.ال.سی) در کتابخانه اصلی دانشگاه ایالتی اوهایو مستقر گردیدند و اولین دفتر کامپیوتری آن در مرکز تحقیقاتی «اوس.سی.یو»<sup>۳</sup> جای گرفت. فردریک کلیگور، اولین مسئول (اوس.ال.سی) بود. در سال ۱۹۹۱ ساختمان مدیریت OCLC توسعه یافت. این شبکه دارای پنج مرکز مدیریت در کلمبو، اوهایو و کتابخانه اصلی دانشگاه ایالتی اوهایو می‌باشد و محل کنونی آن در فرانتس رود ۶۵۶۵ در دوبلین است. اداره (اوس.ال.سی) با اعضای آن است. ساختار مدیریتی آن متشکل از اعضای عمومی، شورای اعضا و هیأت امناء است. (اوس.ال.سی) از یک سیستم کامپیوتری منطقه‌ای برای ۴۵ کالج اوهایو شروع شد و در نهایت به یک شبکه بین‌المللی تبدیل گردید. در سال ۱۹۷۷ اعضای (اوس.ال.سی) از اوهایو، با عضویت کتابخانه‌های خارج از اوهایو موافقت کردند و به این ترتیب، (اوس.ال.سی) به اوس.ال.سی.اینک تغییر نام داد و امروزه انواع مختلف کتابخانه‌ها از سراسر دنیا از خدمات (اوس.ال.سی) استفاده می‌نمایند. به دنبال آن شورای استفاده‌کنندگان (اوس.ال.سی) تشکیل گردید

فهرست آنلاین شروع به کار کرد و در سال مالی ۱۹۷۲، کتابخانه‌ها بیش از ۳۳۰۰۰۰ عنوان را فهرست کردند. در سال ۱۹۷۴ (اوال.یوس.سی) به یک میلیون رکورد، در ۷۶/۱۹۷۵ به دو میلیون، سال بعد به سه میلیون و در دسامبر ۱۹۸۰ به هفت میلیون رکورد کتابشناختی رسید (جردان، ۱۹۹۸). در ۱۹۹۴ این پایگاه شامل بیش از ۲۹ میلیون رکورد بود و نرخ افزایش رکورد آن در هفته ۲۲۰۰۰ رکورد می‌باشد (کروک، ۱۹۹۵).

در ژانویه ۱۹۹۹، (اوس.ال.سی) یک پروژه فوری اولیه تحت عنوان (کرک) یا فهرست منابع پیوسته تعاونی را ارائه نمود. هدف این پروژه که مبتنی بر (دی.سی) یا استاندارد هسته فراداده‌ای در دوبلین بود، توسعه یک فهرست منابع در محیط وب بود. در ژانویه سال ۱۹۹۹ مرکز کامپیوتری کتابخانه اوهایو، پروژه "کورک" را آغاز نمود تا یک پایگاه اطلاعاتی جامع برای منابع پیوسته از طریق تعاونی بین‌المللی با کتابخانه‌ها ایجاد کند.

#### پایگاه‌های اطلاعاتی و سرویس‌های (اوس.ال.سی)

(اوس.ال.سی) به منظور تامین نیازهای انواع کتابخانه‌ها، طیف وسیعی از خدمات را ارائه می‌دهد که شامل ابزار فهرست‌نویسی، پایگاه‌های اطلاعاتی مرجع و خدمات جستجوی پیوسته، ابزار اشتراک منابع، خدمات حفاظت و رده‌بندی دهدهی دیوئی می‌باشد. همچنین امکان ارتباط با فروشندگان متعدد پایگاه‌های اطلاعاتی را میسر می‌سازد.

بطور کلی پایگاه‌های اطلاعاتی (اوس.ال.سی) شامل WorldCat, ArticleFirst, Electronic Collection Online, NetFirst, PAIS International, PaperFirst, Proceedings First, OCLC Union Lists of Periodicals می‌باشد. در اینجا به دو پایگاه‌ها یعنی "وورلدکات" و فهرستگان پایگاه اطلاعاتی مجلات در (اوس.ال.سی) یا بطور مختصر اشاره می‌شود.

در مرکز خدمات (اوس.ال.سی)، پایگاه اطلاعاتی "وورلدکات" قرار دارد که بیشترین استفاده را در آموزش عالی دارد و سابقه دانش بشری را از ۴۰۰۰ سال گذشته

کارهائی که (اوس.ال.سی) انجام می‌دهد شامل خدمات فهرست‌نویسی، جستجو، امانت بین کتابخانه‌ای و فهرستگان ادواریها می‌باشد. بطور کلی می‌توان گفت که در سه زمینه خدمات مرجع و نشر الکترونیکی، توسعه فهرست‌نویسی، خدمات اشتراک منابع و توسعه بین‌المللی پیشرفت داشته است.

#### شبکه و خدمات کتابشناختی (اوس.ال.سی)

شبکه و خدمات کامپیوتری آن با بیش از ۴۱۰۰۰ کتابخانه در ۸۲ کشور و قلمرو ارتباط برقرار می‌نماید. (اوس.ال.سی) از یک پایگاه کتابشناختی به نام "اپیک" نگهداری می‌کند. در سال ۱۹۹۱ این سیستم شامل ۱۶ میلیون کتاب، یک میلیون سریال و ۴۵۰۰۰۰ رسانه سمعی بصری و همچنین ۵۰۰۰۰۰ گزارش بوده است. و تقریباً دو میلیون رکورد سالانه به آن اضافه می‌شود (کاتز، ۱۹۷۴) بطوریکه هم اکنون این پایگاه حاوی بیش از ۳۱ میلیون رکورد در فرمت‌های مختلف کتاب، فیلم، گزارش یا تک‌نگاشت از کتابخانه‌های عضو می‌باشد.

(اپیک) مداخل متعددی از جمله دستیابی موضوعی و کلیدواژه‌ای و جستجوی بولین را فراهم می‌کند. دستیابی موضوعی فقط از سال ۱۹۹۰ میسر بود. نرم‌افزار اپیک علاوه بر رکوردهای اوس.ال.سی امکان جستجوی پایگاه‌های دیگر را از (اریک) و (ان.تی.آی.اس) فراهم می‌سازد. سیستم اپیک هر ساله پایگاه‌های اطلاعاتی جدیدی به پایگاه‌های خود می‌افزاید. به این ترتیب (اوس.ال.سی) بزرگترین منبع اطلاعاتی جهانی است. نواریهای مارک در قلب سیستم (اوس.ال.سی) قرار دارد و نیز شامل فهرست‌نویسی اصلی مهم اعضا و موجودی همه کتابخانه‌های دولتی است.

فهرستگان پیوسته (اوس.ال.سی)، از نظر تعداد کتابخانه‌های عضو و از نظر عمق و وسعت پوشش، پایگاه کتابشناختی پیشگام در دنیاست. تا قبل از سال ۱۹۷۰ اکثر فعالیت‌های (اوس.ال.سی) مربوط به تحقق در مورد طرح و عملکرد سیستم کامپیوتری فهرست آنلاین بود اما پس از آن تعداد ۴۴۰۷۱۱ فهرست‌برگه آف لاین را نیز تولید کرد (پرومن‌شنکل، ۲۰۰۰). در سال ۱۹۷۱ سیستم

یکی از مهمترین سرویس‌های (اوس.ال.سی)، سرویس امانت بین کتابخانه‌ای است که در سال ۱۹۷۹ معرفی گردید و از آن زمان با استفاده از این سرویس بیش از ۱۴ میلیون امانت بین ۶۷۰۰ کتابخانه سراسر جهان انجام گرفت. یکی دیگر از این سرویس‌ها، خدمات جستجوی اولیه می‌باشد که به عنوان یک ابزار مرجع در سال ۱۹۹۱ معرفی شد و اکنون در ۱۹۲۴۶ کتابخانه استفاده می‌شود. وظیفه آن ارائه خدمات در زمینه مجلات الکترونیکی به کتابخانه‌ها و تأمین نیازهای استفاده‌کنندگان راه دور است. این سرویس با ایجاد یک ذخیره اطلاعات و فراهم نمودن امکان اشتراک کتابخانه‌ها (بطور مستقیم از طریق ناشر) در مجموعه چاپی مجلات علمی، نقش خود را در توسعه مجموعه ایفا می‌نماید. مقالات مجلات در (اوس.ال.سی) بطور الکترونیک ذخیره می‌شود و کتابخانه‌ها می‌توانند از طریق وب، با استفاده از یک مرورگر وب اشتراکی و از طریق رابط پشتیبانی‌کننده از جستجو در مجلات، به آن‌ها دست یابند. از طریق خدمات جستجوی اولیه استفاده‌کنندگان می‌توانند به ۷۰ پایگاه اطلاعاتی مشهور و شناخته شده دسترسی پیدا کنند. در حال حاضر این سرویس دارای صفحات جستجو و گرافیک‌های جدیدی است که انجام جستجوهای اساسی و پیشرفته را برای استفاده‌کنندگان به مراتب راحتتر کرده است (هیزل، ۱۹۹۶).

یکی دیگر از سرویس‌های (اوس.ال.سی)، سرویس همکاری کتابخانه‌ای مرکز خدمات غرب است (فور، ۲۰۰۰). این سرویس به منظور استفاده از قدرت جمعی ۱۷۰۰ کتابخانه عضو بر سه نوع خدمات متمرکز است: فراهم‌آوری امکاناتی برای ایجاد پایگاه اطلاعاتی؛ حفاظت و رقیمی کردن منابع؛ و ایجاد فرصت‌های آموزش. ایجاد پایگاه اطلاعاتی، تخفیف‌های قابل توجهی از لحاظ مشترک شدن در منابع الکترونیکی به همراه دارد. حفاظت و رقیمی کردن نیز به اصلاح دستیابی به اطلاعات کمک می‌کند و همچنین به اعضا در نگهداری مجموعه کمک می‌نماید. (اوس.ال.سی) همچنین با کنسرسیوم‌های رسمی و غیررسمی کتابخانه‌ها، موزه‌ها و

تاکنون به ۴۰۰ زبان دربرمی‌گیرد. "ووردکات" بزرگترین پایگاه اطلاعاتی کتابشناختی در جهان است که حاوی بیش از ۴۶ میلیون رکورد فهرست‌نویسی از کتابخانه‌های سراسر جهان بوده و هر پانزده ثانیه یک رکورد جدید به آن اضافه می‌گردد (پرومن‌شنکل، ۲۰۰۰). این رکوردها شامل کتاب، نسخ دستی، نوارهای ویدئویی و صوتی، نشریات ادواری، روزنامه‌ها، نقشه‌ها، و برنامه‌های کامپیوتری می‌باشد. به علاوه محصولات میکروفیلم و میکروفیش را به شکل نوار یا فرمت کارتی فهرست می‌کند. ووردکات طرح‌هایی را برای توسعه مجموعه، فهرست‌نویسی، کنترل مستند، و خدمات گذشته‌نگر ارائه می‌نماید. می‌توان از طریق سیستم‌های محلی به طرو مستقیم به ووردکات دسترسی یافت.

فهرستگان پایگاه اطلاعاتی مجلات در اوس.ال.سی، یکی دیگر از پایگاه‌های اطلاعاتی این شبکه است که شامل میلیون‌ها (ال.دی.آر) می‌باشد و به بیش از ۷۵۰۰۰۰ رکورد کتابشناختی مرتبط شده است. هر (ال.دی.آر) مجموعه اطلاعاتی خاصی از نشریات ادواری را فراهم می‌کند.

سیستم فرعی کنترل نشریات ادواری در (اوس.ال.سی) از سال ۱۹۷۵ مورد استفاده قرار گرفته است و هدف از طراحی آن، فراهم‌آوری امکان مشارکت مؤسسات در کنترل آنلاین نشریات ادواری است. این سیستم دارای چهار جزء است: کنترل، پیگیری، کنترل صحافی و فهرستگان، تا سال ۱۹۸۰ فقط جزء کنترل آن کار می‌کرد اما کتابخانه‌ها می‌توانستند از سیستم کنترل محدود ادواری‌ها نیز استفاده کنند. پردازش یک نشریه ادواری جدید و یا تغییر یک عنوان، یکی از مشکل‌ترین کارهایی است که سیستم فرعی کنترل ادواری‌ها در (اوس.ال.سی) باید انجام دهد. مبنای فرآیند کنترل خودکار سیستم فرعی کنترل ادواری‌ها در (اوس.ال.سی)، کد بسامد یک رکورد داده‌ای محلی است. اما این کدها همیشه روزآمد نمی‌شوند. حل مسائل مربوط به تغییر بسامد، منوط به استفاه از جزء پیگیری کنترل ادواری‌هاست (رافتون، ۱۹۸۲).

(multidrop) متداولترین روش دستیابی باقی ماند. در سال ۱۹۹۱، (ا.وسی.ال.سی) یک شبکه جدید X.25 را نصب کرد که نه تنها سریع، قابل اعتماد و از نظر هزینه، سودمند بود بلکه می‌توانست با نیازهای فزاینده خدمات کتابشناختی (ا.وسی.ال.سی) سازگار باشد. در سال ۱۹۹۵، (ا.وسی.ال.سی) شروع به فراهم آوردن دستیابی اینترنتی برای خدمات فهرست‌نویسی و خدمات اشتراک منابع خود نمود (جردان، ۱۹۹۸). استفاده کننده می‌تواند به طور خودکار به (ا.وسی.ال.سی) وصل شود و پایگاه‌های اطلاعاتی انتخابی خود را جستجو نماید. همچنین می‌تواند به طور مستقیم و یا از طریق شبکه‌های منطقه‌ای مثل دانشگاه ایالتی شبکه کتابخانه‌ای (ا.وسی.ال.سی) نیویورک یا "نلی‌نت" به (ا.وسی.ال.سی) دسترسی یابد. هزینه‌های ورودی در هر یک از این سیستم‌ها متفاوت است. هزینه‌های (ا.وسی.ال.سی) شامل اسنادهای آنلاین و نیز تجهیزات، نگهداری و هزینه‌های ارتباط راه دور هستند (کاتز، ۱۹۷۴).

(ا.وسی.ال.سی) بر روی دیسک‌های فشرده نیز قابل دسترسی می‌باشد. مجموعه دیسک‌های فشرده (کت) CD450 حاوی زیرمجموعه‌هایی از رکوردهای پایگاه اطلاعاتی اصلی می‌باشد که بطور فصلی روزآمد می‌شوند. این دیسک‌ها در پنج زمینه موضوعی کشاورزی، آموزش، هنر، علوم انسانی، علوم و تکنولوژی به فروش می‌رسند. هر مجموعه، علاوه بر افزایش نمایه‌ها، دستیابی به رکوردهای (ا.وسی.ال.سی) در این زمینه‌های موضوعی را فراهم می‌کنند. مثلاً بسته نرم‌افزاری آموزشی شامل رکوردهای (ا.وسی.ال.سی) در آموزش و نیز دیسک‌های جاری و گذشته‌نگر اریک می‌باشد.

### پروژه‌های (ا.وسی.ال.سی)

(ا.وسی.ال.سی) دارای پروژه‌هایی است که عبارتند از: Kilory, Scorpion, Marc, Xml, (آر.دی.اف)، Dublin, Kilory, WordSmith به طور کلی اینترنت را مورد جستجو قرار می‌دهد و با پایگاه‌های اطلاعاتی تمام متن و پایگاه‌های اطلاعاتی فراداده‌ای از منابع اینترنتی ارتباط برقرار می‌کند. Iteract که در

آرشیوها در تعیین روش‌های حفاظت مشارکت دارد. بخش حفاظت از منابع، اقدام به ارائه خدمات میکروفیلمی و رقمی در غرب می‌نماید و نیز با شبکه‌های منطقه‌ای و سایر سازمان‌ها برای تهیه خدمات مشاوره‌ای و آموزشی مرتبط همکاری می‌نماید. بطور کلی می‌توان گفت مهمترین فعالیت‌های اصلی مرکز خدمات غرب، عبارتند از: برگزاری کارگاه‌های آموزشی و به نمایش گذاشتن محصولات و خدمات (ا.وسی.ال.سی)، توسعه فعالیت‌های مربوط به آموزش فهرست‌نویسی، توسعه برنامه‌های مدیریت کتابخانه، خدمات مرجع، حفاظت و رقمی‌سازی (پرومن‌شنکل، ۲۰۰۰).

### (ا.وسی.ال.سی)، کتابخانه‌ها و اینترنت

محیط وب و اینترنت، مکانیزم دیگری برای تحول مدرک در راستای هدف عمومی (ا.وسی.ال.سی) یعنی افزایش دستیابی به اطلاعات جهانی و کاهش هزینه‌های اطلاعاتی است در اختیار (ا.وسی.ال.سی) قرار می‌دهد. شبکه اینترنت روش‌های جدید و مهیجی در زمینه خدمات فنی فراروی (ا.وسی.ال.سی) و کتابخانه‌های عضو قرار می‌دهد. کتابخانه‌های عضو (ا.وسی.ال.سی) به سرعت در حال افزایش استفاده از خدمات وب و اینترنت می‌باشند. در حال حاضر تقریباً ۸۵٪ از کاربرد جستجوی اولیه و ۱۸٪ از فهرست‌نویسی پریزم و اشتراک منابع بر روی اینترنت موجود است. (ا.وسی.ال.سی) از طریق <http://www.oclc.org> در روی وب قابل دسترسی است. با افزایش تعداد کتابخانه‌ها، کنسرسیوم‌ها و شرکت‌هایی که به اینترنت وصل می‌شوند، این شبکه جهانی با محدودیت‌های فنی روبرو می‌شود (لایبریز، ۲۰۰۲). در همین زمینه مهندسين شبکه (ا.وسی.ال.سی) و فراهم‌آوردندگان خدمات اینترنت جهت رفع مسائل ناشی از افزایش ترافیک اینترنت که بر قابلیت دسترسی و زمان پاسخگوئی تأثیرگذار است، با یکدیگر همکاری می‌نمایند.

### دسترسی به (ا.وسی.ال.سی)

(ا.وسی.ال.سی) در ابتدای تشکیل خود، در سال ۱۹۷۱، یک امکان دستیابی ویژه به نام (multidrop dedicated) و در ۱۹۷۴، یک امکان دستیابی تلفنی (dial-access) فراهم نمود. بیست سال بعد

(گودبی، ۱۹۹۸). چند خروجی نمایه در شکل ۱ نشان داده شده است.

Have	havey	haw
Havei	havice	hawing
Havel	havill	hawed
Haven	havilland	hawker

شکل ۱. نمایش قسمتی از یک نمایه تکواژه‌ای

مدخل نمایه، اسم کامل یک مفهوم مشخص را در متن تعیین می‌کند مثلاً عباراتی که در شکل‌های ۲ و ۳ آمده است.

Air pollutants	air power
Air pollution	air pressure
Air pollution	air quality

شکل ۲. عبارات اسمی در یک پایگاه اطلاعاتی در محیط وب

Yevette	Abel Tasman
Haskins	National Park
Yevgeny	Abercorn
KafeInikov	Street

شکل ۳. اسامی اشخاص و اسامی جغرافیائی در یک پایگاه اطلاعاتی در محیط وب

بطور کلی مجموعه‌ای از مداخل نمایه چکیده محتوایی یک پایگاه اطلاعاتی را نشان می‌دهد و سهولت دستیابی به انواع چکیده‌ها را فراهم می‌سازد. به عنوان مثال متداولترین اسامی اشخاص در پایگاه اطلاعاتی مقالات خبری استخراج شده از وب در شکل ۴ نشان داده شده است.

Bob Dole	L Gore
Greg Noble	ennis
Gerge Washington	George Bush

شکل ۴. اسامی افراد در پایگاه اطلاعاتی مقالات خبری

از طرف دیگر نمایه‌های عبارتی دربرگیرنده موضوعاتی مثل male breast cancer است که هنوز در سیستم‌های رده‌بندی‌ای مثل دیوئی جانی برای آن در نظر گرفته نشده است.

با توجه به موارد بالا، این نوع نمایه‌سازی برای پایگاه اطلاعاتی خیلی بزرگ تفاوت بسیاری با نمایه‌سازی مدارک تمام متن یا غیر ساخت یافته دارد. این موضوع

سال ۱۹۹۳ شروع و در سال ۱۹۹۶ تکمیل گردید. عبارت است از یک سیستم بررسی تحقیقاتی (اوس.ال.سی) از نظر ماهیت و اندازه منابع اینترنتی و تأثیر بالقوه آن‌ها بر عملیات کتابخانه‌ای Mantis به منظور سازماندهی و دستیابی به منابع الکترونیکی، استفاده و جامعیت سیستم‌های Kilory, Scorpion, Marc, Xml, (آر.دی.اف)، Dublin Core و WordSmith را مورد تحقیق و بررسی قرار می‌دهد. اسکورپیون پروژه جدیدی است که به بررسی نمایه‌سازی و فهرست‌نویسی منابع الکترونیکی می‌پردازد. با توجه به افزایش اطلاعات موجود در وب، و در نتیجه افزایش مدت زمان لازم برای یافتن اطلاعات مورد نیاز و نیز افزایش هزینه دستیابی به آن، پروژه اسکورپیون در نظر دارد نمایه‌سازی و فهرست‌نویسی را با یکدیگر ترکیب کند و مخصوصاً ابزاری برای تشخیص خودکار موضوع براساس طرح‌های شناخته شده رده‌بندی دیوئی تهیه نماید. در صورتی که سرعنوان‌های موضوعی یا دامنه‌های مفهومی را بتوان بطور خودکار به مدارک الکترونیک تخصیص داد، می‌توان در زمینه ایجاد ابزار فیلترکننده‌ای برای جستجو اقدام نمود (شیفر، ۱۹۹۶). پروژه وورداسمیت، سیستم نمایه‌سازی اوس.ال.سی است که محور اصلی این مقاله را تشکیل می‌دهد.

## پروژه "دبلیو.اس" در پایگاه کتابشناختی (اوس.ال.سی)

یکی از اهداف کوتاه مدت پروژه (اوس.ال.سی)، توسعه نرم‌افزاری است که بتواند عبارات را بطور خودکار تشخیص داده و این عبارات را هوشمندانه در مدارک تمام متن سازماندهی کند و به این ترتیب نمایه‌های مفیدتری برای پایگاه‌های اطلاعاتی ایجاد گردد. به همین منظور با استفاده از روش‌های توسعه‌یافته در زبان‌شناسی محاسباتی و نیز استفاده از نرم‌افزار NetOwl امکان تشخیص و مقوله‌بندی اسامی خاص در متن غیرساختاری و نیز تهیه لیست‌های مقوله‌بندی شده مناسب از اسامی اشخاص و اسامی جغرافیائی فراهم آمد

ماشینی انجام گرفت (گریش‌مان، ۱۹۸۴). بیش از سه دهه گذشته، این نوع پردازش مجموعه‌ای از مدل‌های آماری، زبانشناسی و معنایی را مورد توجه قرار داده است (دزکوچ، ۱۹۸۶). یکی از ویژگی‌های (ان.ال.پی) این است که می‌توان ساختارهای زبانشناسی متفاوتی برای انتقال یک حقیقت بکار برد. در نتیجه برای کنترل متون زبان طبیعی حتی در یک دامنه محدود، به قوانین زیادی نیاز است. این قوانین باید سازماندهی شود و برای انجام این کار، لازم است ابتدا «ساختار اطلاعات» را در همان حیطه پردازش متون مشخص نمود. به این ترتیب اقدام به رده‌بندی و تعیین نحوه ترکیب این حقایق برای تشکیل ساختارهای بزرگتر نمود.

قالب‌ها و شکل‌های متعددی برای اشراف بر دامنه اطلاعات ایجاد شده است. بعضی از این قالب‌ها تحت عنوان «سیستم‌های مبتنی بر چارچوب» یا «frame-based systems» نامیده شده است. یک چارچوب یا عبارت است از یک ساختار داده‌ای برای نگهداری اطلاعاتی در مورد نمونه‌ای از یک رده خاص موضوعات. یا قالبی که مخصوصاً برای تحلیل متون علمی و فنی ایجاد شده است «فرمت یا قالب اطلاعاتی» نام دارد. تحلیل نحوی بیشترین زمینه‌ای است که در زبانشناسی محاسباتی مورد مطالعه قرار گرفته و نسبتاً موفق بوده است (گریش‌مان، ۱۹۸۴). زبانشناسی محاسباتی حوزه‌ای است که به سرعت در حال رشد است و بسیاری از مسائل متفاوت در پردازش زبان را مورد توجه قرار می‌دهد. تحلیلگرهای زیادی دارد که برای کارهای ساده پرسش و پاسخ تعداد کمی از آن‌ها کفایت می‌کند. اما در مقایسه با توانایی انسان در درک زبان، هیچ یک از این تحلیلگرها عالی نیستند. حتی اگر بتوانند تطابق‌های ورودی برای الگوهای مورد انتظار نحوی و معنایی را بخوبی انجام دهند، لیکن از عهده کنترل ورودی که قادر به تأمین این انتظارات نیست بر نمی‌آیند. انسان انعطاف‌پذیرتر است و می‌تواند در صورت عدم رعایت هنجارهای نحوی، از کلیدهای معنایی و نیز کلیدهای نحوی استفاده نماید. به هرحال هم انسان و هم کامپیوتر می‌توانند اطلاعات را کنترل کنند. مثلاً هنگامی که می‌گوییم Mary ate

یکی از اهداف اساسی پروژه دلبیواس برای ثبت فنون پردازش زبان طبیعی به منظور قابل دسترس ساختن اطلاعات جهانی است (گودبی، ۱۹۹۸).

در سال ۱۹۹۶ دفتر تحقیقات، یک برنامه تحقیقاتی را تحت عنوان پروژه دلبیواس پردازش زبان طبیعی آغاز کرد. هدف اولین مرحله این پروژه، توسعه روش‌های اساسی برای تشخیص اصطلاحات مربوط به موضوعات فرهنگی بود. نتایج را می‌توان در نمایه‌ها، اصطلاحنامه‌ها و سیستم‌های رده‌بندی که منجر به افزایش دستیابی به پایگاه‌های اطلاعاتی (اوس.ال.سی) می‌شوند گنجانده؛ چون تمرکز بر روی تشخیص و تحلیل اصطلاحات است. دلبیواس دارای پروژه دیگری به نام وورداسمیت توسعه یافته است. این پروژه با بکارگیری زبانشناسی محاسباتی، یک سری فیلترهای آماری را تکمیل کرده و به بررسی امکان‌سنجی استخراج مستقیم اصطلاحات موضوعی از متن خام می‌پردازد. و به این ترتیب با استفاده از فن مشابهی اصطلاحات تزاروسی را به طور خودکار تولید می‌کند (خان، ۲۰۰۲). از جنبه عملی‌تر، توسعه‌ای که اخیراً در فناوری «LEXIS/NEXIS SmartIndexing» ایجاد شده، خصوصیات واژگان کنترل شده را با الگوریتم نمایه‌سازی ترکیب می‌کند تا با مواردی مانند بسامد، وزن‌دهی، و تعیین محل مدرک در نمایه‌سازی مجموعه‌های خبری LEXIS/NEXIS ارتباط پیدا کند.

قبل از ادامه بحث پیرامون ساختار و مدل‌های نمایه‌سازی دلبیواس لازم است درباره پردازش زبان طبیعی که در این نمایه مورد استفاده قرار گرفته است توضیحاتی ارائه گردد.

### پردازش زبان طبیعی

پردازش زبان طبیعی یا (ان.ال.پی) در ارتباط با توسعه برنامه‌های کامپیوتری است که می‌تواند زبان طبیعی را تجزیه و تحلیل کرده و در زمینه اطلاعات مندرج در متن یا گفتار، عملکرد مناسبی داشته باشد. (ان.ال.پی) تاریخ طولانی دارد و شروع آن همزمان با تلاش‌هایی بود که در سال ۱۹۵۰ در زمینه ترجمه



زندگی واقعی استفاده می‌کند و به رفع ابهام معنایی کمک می‌نماید.

پردازش زبان طبیعی یک جزء جدائی‌ناپذیر از کاربردهای هوش مصنوعی است. هوش مصنوعی حوزه وسیعی از علم کامپیوتری می‌باشد که مربوط به رفتار هوشمند برنامه‌های کامپیوتری است و شامل سؤال و جواب، ترجمه ماشینی، یادگیری ماشینی، سیستم‌های ماهر و جلودارهای هوشمند برای سیستم‌های پایگاه اطلاعاتی-مدیریتی می‌باشد. مشخصه (ان.ال.پی) در هوش مصنوعی، زبان‌شناسی قوی و پیچیده، و توانایی‌های منطقی یعنی استنتاجات قیاسی و استقرائی آن است. به هر حال در هر دو زمینه هوش مصنوعی و (آی.آر) نیاز به یک تئوری واحد درباره زبان و منطق برای کنترل زبان طبیعی و تحلیل معنایی خودکار و ساختاربندی حجم عظیمی از متن زبان طبیعی برای بازیابی اطلاعات و هوشمند ساختن دانش جدید وجود دارد. از نقطه نظر پردازش زبان طبیعی و سیستم‌های (آی.آر)، سیستم‌های (دی.بی.ام.اس) و سیستم‌های هوش مصنوعی یا نشان‌دهنده پیشرفت ساختاربندی و پیچیده‌تر شدن عملکردهای پردازش اطلاعات است (دزکچ، ۱۹۸۶). به طور کلی سیستم‌های موجود بازیابی اطلاعات، سیستم‌هایی اساساً منفعل هستند که توسط جستجوکنندگان انسانی هوشمند استفاده می‌گردند. سیستم نمایه‌سازی دبلیو.اس، یک نظام نمایه‌سازی خودکار براساس پردازش زبان طبیعی و برنامه‌های هوشمند کامپیوتری استوار است.

### ساختار نمایه‌سازی دبلیو.اس

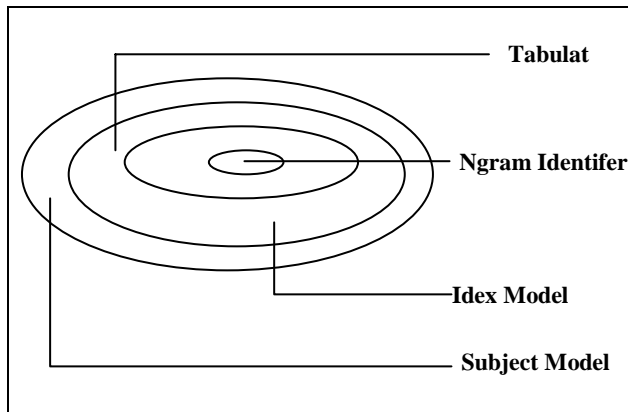
نمایه‌سازی (دبلیو.اس)، سیستمی خودکار است که بر مبنای زبان‌شناسی محاسباتی و تئوری رده‌بندی استوار است. هدف این سیستم، اصلاح دستیابی استفاده‌کننده به مجموعه الکترونیکی از طریق توسعه روش‌های مؤثر در تشخیص و سازماندهی کلیدواژه‌های مهم در محتوای یک مدرک است. مبنای کار در این نمایه‌سازی، تشخیص خودکار مفاهیم در متن

pudding علاوه بر معنای آن، اکثر مردم معنای Mary pudding ate را نیز از آن درک می‌کنند حتی اگر خلاف الگوهای نحوی معمولی باشد. همچنین از عبارت "Mary ate frapple" حدس خواهند زد که frapple یک نوع غذاست حتی اگر قبلاً این کلمه را نشنیده باشند. برای رفع ابهام از این سؤال، نیاز به اطلاعات زمینه‌ای است. باید به موانع و محدودیت‌های محلی توجه کرد و همچنین رده‌های معنایی اسم، فعل و صفت را تعریف و سپس محدودیت‌هایی که از نظر با هم قرار گرفتن این رده‌ها وجود دارد را مشخص نمود.

حوزه‌های زبان طبیعی دارای دو کاربرد اصلی هستند: «ماهیت طبیعی بازیابی پایگاه اطلاعاتی» و «ماهیت غیرتعاملی ساختاربندی متن». پردازش زبان طبیعی، ثبت برخی از انواع اطلاعات را از طریق مکالمه انسان-ماشین، مؤثرتر می‌کند.

### سطوح پردازش زبان طبیعی

بررسی امکانات (آی.آر) برای (ان.ال.پی) از نظر سطوح پردازش زبان‌شناسی محاسباتی عبارتند از: (۱) سطح فونتیکی یا آوایی مربوط به صداهای گفتار؛ ۲ سطح مورفولوژیکی مربوط به پردازش شکل‌ها و بخش‌های قابل تشخیص هر کلمه؛ ۳ سطح واژگانی مربوط به عملکرد کلمات؛ ۴ سطح نحوی در تحلیل زبان مربوط به تشخیص واحدهای ساختاری مثل عبارات اسمی. محققان زبان‌شناسی محاسباتی و هوش مصنوعی ابزار تجزیه‌گر خودکار و پیشرفته‌ای برای استفاده در ان.ال.پی ایجاد کرده‌اند. از جمله موجودیت‌های نحوی مثل عبارات اسمی در شکل سرعنوان‌های موضوعی، امکانات کنترل آزاد متن مثل عملکردهای هم‌جواری و تطابق الگو، تسهیلاتی برای محدود کردن جستجو به مرزهای متنی خاص مثل عنوان که به منزله ساختارهای مهم نحوی در (آی.آر) به کار می‌رود. (۵) سطح معنایی تحلیل مربوط به استفاده از دانش متنی برای نشان دادن معنای متن زبان طبیعی؛ (۶) سطح عملی تحلیل زبان از اطلاعات مربوط به موضوعات



شکل ۵. فیلترهای چهارگانه در سیستم نمایه‌سازی وورداسمیت استفاده از آمار همبندی در متنی که درباره نمایه‌سازی خودکار باشد، نشان می‌دهد که احتمال این که کلمه indexing بعد از کلمه automatic بیاید بسیار زیاد است. فایلهائی که به وسیله جدول‌ساز ایجاد می‌شوند، مبنائی هستند برای همه پردازش‌های بعدی.

فیلتر دوم، یعنی معرفه ان‌گرام یا یک مرحله پیش پردازشی است که حدود ۲۰٪ از متن خام که دارای بیشترین احتمال از نظر داشتن واژگان قابل نمایه شدن باشد، را مشخص می‌نماید. این معرفه، توالی‌های شکسته نشده کلمات همبند را با استفاده از متن خام و فایل بیگرمهای همبند، به عنوان ورودی می‌یابد. این کلمات ممکن است طولانی‌تر از دو کلمه باشند. ورودی ممکن است شامل توصیفگرهای موضوعی مفیدی مثل "WordSmith Indexing System" باشد. از آنجا که در مورد خروجی مطلوب، پیش فرضهای اندکی وجود دارد لذا معرفه "ان‌گرام" بیشتر رشته‌های متنی طولانی‌تری که احتمالاً مبنی قراردادهای انشائی در یک حوزه موضوعی بوده و ساختارشان قابل پیشگوئی نباشد را دربر می‌گیرند. مثلاً ان‌گرام‌های مجموعه‌ای از اخبار سیاسی، دربردارنده طیف وسیعتری از مسائل می‌باشند.

مسائلی مانند حرص، تباهی، تحول در رهبری و غیره. فیلتر دوم، مدل نمایه است و هدف آن ایجاد واژگان مفیدی است که قابل نمایه شدن باشند. برای انجام این کار، خروجی معرفه ان‌گرام را دستکاری و مرتب کرده، آن‌ها را تبدیل به عبارات کوتاهی که دارای ساختار استاندارد باشد، می‌نماید. از آنجا که دامنه بالقوه وسیعی از واژگان مفید وجود دارد، مدل نمایه بسته به خصوصیات

ماشین‌خوان است. برای تشخیص کلمات و عباراتی که به بهترین نحو مبنی موضوع مدرک باشند از تئوری‌های زبان‌شناسی و اطلاع‌رسانی استفاده می‌گردد. به این ترتیب دائماً برای پایگاه‌های اطلاعاتی، اصطلاحات موضوعی تهیه می‌شود. قسمتی از نمایه در ذیل آورده شده است:

Ad
Ad hominem
Ad Limina
Army officer
Army officer crops
Asian
Asian republics
Attack
Attack ads
Balance
Balance sheet
Balance sheets
Balanced budget
Balanced budget amendment

نمونه‌ای از نمایه وورداسمیت

در طراحی سیستم نمایه‌سازی (دبلیو.اس) چهارمدوله برای مشخص کردن واژگان موضوعی وجود دارد که در شکل ۵ نشان داده شده است. هر مدوله الزاماً خصوصیت فیلتری را دراد که جستجو را به توصیفگرهای موضوعی بالقوه در متن ماشین‌خوان نامحدود می‌سازد. هر چه تعداد فیلترها بیشتر باشد، پردازش پیچیده‌تر و متمرکزتر می‌شود.

فیلتر اول جدول‌ساز است که وظیفه آن ایجاد فایلهای داده‌ای می‌باشد. با توجه به طرح لایه‌ای این سیستم، متمرکزترین پردازش ممکن است به میزان ۰.۰۵٪ فایل داده‌ای اصلی محدود شود. فایلهای شامل: بسامد واژه‌ها، بسامد بیگرمهای متوالی و بسامد بیگرمهای متوالی که همبند آن‌ها با یکدیگر خیلی کم است. مثلاً اگر مقاله‌ای در زمینه نمایه‌سازی دبلیو.اس در اختیار جدول‌ساز قرار گیرد، بیگرمهای متوالی عبارتند از: The WordSmith, WordSmith Indexing & indexing System. و این روند به همین ترتیب ادامه می‌یابد تا همه جفت کلمات نزدیک به هم مشخص و شمارش شوند. بیگرمهای همبند حدود ۲۵٪ از بیگرمها را نشان می‌دهند. این امر بوسیله یک آمار همبندی ساده مشخص می‌گردد (گودبی، ۱۹۹۸).

مجموعه خاصی از مدارک استخراج شده باشند، خوشه‌هایی را تشکیل می‌دهند که مبتنی بر شباهت نحوی می‌باشد. در مجموعه مدارکی که در مورد نرم‌افزار کاربردی کامپیوتری باشد، کلمه "file"، کلمه‌ای متداول است زیرا نه تنها به عنوان یک اصطلاح مهم در متن ظاهر می‌شود بلکه کلمه‌ای است که در رأس عبارات اسمی مثل (پی.دی.اف) File, HTML file و غیره قرار می‌گیرد. همچنین در مورد مدارکی که در زمینه معماری می‌باشد، کلمه "design" یک کلمه متداول است و در عباراتی مثل architectural design, urban design و غیره ظاهر می‌شود. در زبانهای غیرانگلیسی، ممکن است سرعنوان عبارت اسمی، مستقیم‌ترین اسم نباشد؛ اما بدون توجه به محل ظاهر شدن آن کلید قابل اعتمادی برای تشخیص واژگان موضوعی غنی و پرمحتوا باشد (گودی، ۱۹۹۸) می‌گوید آکه‌گاوا و واخ‌هولدر که از متخصصین زبان‌شناسی محاسباتی هستند، توالی‌های این مشاهدات را مورد بررسی قرار می‌دهند.

به منظور تشریح خروجی ایجاد شده توسط سیستم نمایه‌سازی (دبلیو.اس)، از همه فیلترهای این سیستم استفاده می‌شود تا عباراتی را از دو حیطة موضوعی متفاوت یعنی ستاره‌شناسی و مطالعات پژوهشی به دست آوریم. متداولترین سرعنوان‌های عبارات اسمی مربوط به هر مجموعه از متون در زمینه مقالات این دو حوزه به شرح زیر است:

Head	Sample Phrase
Framework (s)	Resource Descriptin Framework, WarWick Frmework
Library (ies)	Alexandria Digital Library, Digital Libraries, Cornel University Library
Qualifier	Dc Qualifiers, Default Qualifiers, Dublin Core Qualifiers

جدول ۱. سرعنوانهای استخراج شده از حوزه مطالعات پژوهشی

Head	Sample Phrase
Material	organic material, circumsteller material
Cluster (s)	tapezieum cluster, globular cluster, beehive cluster
Way	milky way, summer milky way, winter milky way
Eclips (s)	parial solar eclips, annual eclips, solar eclips

جدول ۲. سرعنوانهای استخراج شده از حوزه ستاره‌شناسی

متن ورودی و نیاز استفاده‌کننده، دارای پارامترهای بسیار زیادی است. وظیفه این پارامترها، مشخص کردن معیارهای خاصی است که عبارات قابل نمایه شدن می‌توانند تأمین کنند. مدل نمایه همچنین با استفاده از «برچسب جزئی از گفتار» می‌تواند با ایجاد تغییرات ساده در یک فایل شکلی، تئوری‌های بسیاری را در مورد ساختار اصطلاحات نمایه مورد استفاده قرار دهد. مثلاً بطور معمول فرض بر این است که عبارات اسمی نرمال شده را به عنوان توصیفگرهای موضوعی خوب در نظر بگیرند؛ مانند: aerpspace industry, aviation fuel tax, bank fraud, beef crisis شامل توالی اسمی هستند اما عبارات اسمی که به وسیله صفات مشخص می‌شوند نیز می‌توانند توصیفگرهای موضوعی باشند؛ مثل: big labour, biological warefare agent .. بعضی از حوزه‌های موضوعی ممکن است شامل عبارات اسمی زیادی با حروف ربط یا حروف اضافه‌ای باشند؛ مثل: Library of Congress Subject Headings, high crimes. اما همه عبارات نمایه به همین راحتی نیستند. این موارد را می‌توان با استفاده از تعداد آمار همایندی که به وسیله مدوله جدول‌ساز تأمین می‌شود، تشخیص داد. همه این متغیرهای موجود در مدل نمایه را می‌توان دستکاری کرد و واژگان مفیدی که محتوای پایگاه اطلاعاتی را دقیق‌تر از تک‌واژه‌ها منعکس کند ایجاد نمود. بسته به حد و مرزهای تعیین شده در فایل شکلی، مدل نمایه ممکن است ۰.۰۵٪ از فایل متن اصلی را نشان دهد.

مدل موضوعی آخرین مرحله در مشخص کردن واژگان موضوعی است. این مدل در صورتی مورد استفاده قرار می‌گیرد که متن ورودی از یک موضوع محدود گرفته شده باشد یا به عبارت دیگر مربوط به نمایه‌سازی یک متن موضوعی خاص باشد. دستیابی به چنین متنی از طریق وب امکان‌پذیر است. واژگان برگرفته شده از موضوعات محدود، یکسری قواعد اضافی را نشان می‌دهد که با استفاده از چند فرآیند آماری نسبتاً ساده می‌توان آن‌ها را مشخص و مرتب نمود. برای مثال، عباراتی که به منزله خروجی مدل نمایه هستند و همگی آن‌ها از

hole به this galaxy, a galaxy و غیره. اما کلمه hole به حدی مبهم است که نمی‌تواند به تنهایی، به عنوان یک اصطلاح غنی موضوعی در ستاره‌شناسی به کار رود مگر این که به همراه یک توصیفگر بیاید. در حوزه مطالعات پژوهشی نیز (مانند حوزه ستاره‌شناسی)، کلمات مهمی وجود دارد که دارای بار معنایی هستند. مانند واژه‌های software, technology, internet, HTML way, از طرف دیگر واژه‌های خیلی مبهمی مثل milk core و غیره اساساً در عبارات پیچیده‌ای مثل Dublin core way و مشاهده می‌شوند. تک واژه‌های غنی موضوعی به نوبه خود ارزشمندند و به علاوه پایه و اساس مهمی نیز برای ایجاد ساختارهایی به مراتب پیچیده‌تر از عبارات طولانی‌تر می‌باشند.

### نتیجه‌گیری

با توجه به نقش نمایه‌ها در تحقق اهداف اساسی پایگاه‌ها و شبکه‌های کتابشناختی-یعنی افزایش قابلیت دسترس‌پذیری به اطلاعات-به اهمیت ساختار و ماهیت آن‌ها پی می‌بریم. شبکه (ا.و.سی.ال.سی) در بین سایر شبکه‌ها، به عنوان بزرگترین شبکه کتابشناختی، رسالتی مهم برای ایجاد تسهیلات بازیابی منابع اطلاعاتی الکترونیکی و تهیه سیستم نمایه‌سازی قوی و کارآمد که از اثربخشی کافی در محیط وب برخوردار باشد، بر عهده دارد. به همین علت این شبکه با توجه به محیط پویای شبکه و ماهیت اطلاعات وبی، اقدام به تهیه پروژه نمایه‌سازی خود تحت عنوان دبلیو.اس نموده است.

شبکه (ا.و.سی.ال.سی) از روشها و فنون در سیستم نمایه‌سازی خود استفاده نموده است که از قابلیت و توان کافی برای هم‌خوانی با این بزرگترین پایگاه کتابشناختی جهانی برخوردار باشد. به عبارت دیگر ضمن این که به بهترین نحوی نمایانگر محتویات پایگاه اطلاعاتی (ا.و.سی.ال.سی) باشد تسهیلات لازم را نیز برای جستجوی اطلاعات الکترونیکی به طور مؤثر و کارآمد فراهم آورده و رضایت استفاده‌کننده را تأمین نماید. از جمله این فنون می‌توان به استفاده این سیستم از فنون پردازش زبان

### عبارات نمایه‌سازی و ابهام در نمایه‌سازی دبلیو.اس

عبارات استخراج شده از متن مدرک، همیشه بروشنی بیانگر موضوع آن نمی‌باشند. در بیشتر موارد چنین است. هر یک از کلمات تشکیل دهنده عبارات در جداول ۱-۲ به تنهایی مبهم هستند اما هنگامی که با یکدیگر ترکیب می‌شوند، در ارتباط با حیطه موضوعی خود بسیار خاص می‌شوند و مفهوم آن حوزه را به روشنی بیان می‌کنند. مانند عبارت "ring system" که در متون ستاره‌شناسی متداول است. کلمه "system" معانی متعددی دارد اما عبارت "ring system" کاملاً روشن و ارتباط خاصی با حوزه ستاره‌شناسی دارد. کلمه "System" دارای مفاهیم متعددی می‌باشد. این مفاهیم عبارتند از:

1. a procedure or process for obtaining an objective,
2. an organized structure for arranging or classifying,
3. physically chemistry,
4. the living body considered as made up of interdependent components forming a unified whole,
5. a complex of methods or rules governing behavior

در این مدل می‌توان همچنین نتایج مربوط به تأثیرات ابهام را به طور دقیق‌تری مشاهده نمود. مثلاً متون مربوط به ستاره‌شناسی دربرگیرنده عباراتی است که از کلمه galaxy که واژه‌ای نسبتاً روشن و بدون ابهام می‌باشد ساخته می‌شود. مانند Cartwheel galaxy و Cluster galaxy. متون ستاره‌شناسی همچنین شامل عباراتی است که از کلمه Hole که واژه‌ای بسیار مبهم است، ساخته می‌شوند؛ مانند black hole یا grey hole. نمایه‌ای مفید در زمینه اصطلاحات ستاره‌شناسی ممکن است دارای لیستی از انواع galaxy باشد که به طور سلسله‌مراتبی تنظیم شده‌اند. در حالی که ممکن است در مورد واژه hole چندان مفید نیست. Galaxy واژه‌ای مهم در ستاره‌شناسی است که می‌تواند بدون توصیفگر صفتی یا اسمی در عبارات اسمی بیابند؛ مانند

قبل به فرضیه پردازی نمی‌پردازد؛ بنابراین از نظر اصولی می‌تواند اصطلاحات موضوعی غنی را در زبانهای غیرانگلیسی مشخص نماید.

هر زمان که اصطلاحات غنی موضوعی مورد نظر باشد، اصطلاحنامه تهیه شده توسط سیستم (اوس.ال.سی) مورد استفاده قرار می‌گیرد. به علاوه با استفاده از این اصطلاحنامه، امکان تشخیص واژگان مربوط به انواع راهبردهای مروری در مجموعه مدارک موجود در وب وجود دارد. با توجه به ثبات این اصطلاحنامه در زمانهای مختلف، می‌توان به عنوان یک منبع ثابت مثل طرح رده‌بندی یا فرهنگ لغت، به آن مراجعه نمود. همچنین امکان تبیین و استفاده از آن برای طرحهای رده‌بندی موجود مثل دیوئی، با استفاده از روشهای خودکار یا نیمه خودکار وجود دارد.

با توجه به این که در مقایسه با سایر سیستم‌های نمایه‌سازی، با استفاده از این سیستم می‌توان به حجم عظیمی از متن ماشین‌خوان در بسیاری از موضوعات دسترسی یافت، امکان پیشرفت آن وجود دارد. و از طرف دیگر با توجه به افزایش توزیع فنون الکترونیکی در عصر فناوری اطلاعات، و نیز فقدان زمان کافی برای تحلیل و رده‌بندی آنها با استفاده از روشهای سنتی دستی، به چنین پیشرفت‌هایی نیاز است.

طبیعی اشاره نمود. طبق تجربیات متعدد در زمینه آزمون و ارزیابی زبان‌های نمایه‌سازی در طول دهه گذشته، زبان طبیعی به عنوان بهترین زبان نمایه از نظر اثربخشی و کارآئی شناخته شده است (باتاخریا، ۱۹۷۴). پردازش زبان طبیعی بر تئوریهای زبانشناسی، آماری، اطلاع‌رسانی و معنایی تأکید دارد. نمایه‌سازی دلبیواس با استفاده از این تئوریه‌ها و برخورد ظریف و پیچیده با واژه‌ها موفق به طراحی، ساخت و ایجاد سیستم نمایه‌سازی مفید و ارزنده جهت تحقق هدف بنیادی خود برای دستیابی به حجم زیادی از اطلاعات الکترونیکی موجود در وب شده است.

سیستم نمایه‌سازی (اوس.ال.سی) با توجه به یک سری از فیلترهای متن خام، اصطلاحات را مشخص می‌کند که این اصطلاحات از نظر محاسباتی قابل بررسی می‌باشد. این سیستم برای تشخیص واژگان موضوعی دارای ۴ فیلتر می‌باشد: (۱) جدول‌ساز مسئول ایجاد فایل‌های داده‌ای با استفاده از آمارهای همبندی می‌باشد. (۲) فیلتر معرفه ان‌گرام که یک مرحله پیش‌پردازشی است و بیشترین قسمت از متن خام را که ممکن است دارای واژگان قابل نمایه شدن باشد مشخص می‌کند. (۳) مدل نمایه که مسئول ایجاد واژگان مفید قابل نمایه‌شدن است. و (۴) مدل موضوعی که در صورت نمایه‌سازی متن موضوعی خاص مورد استفاده قرار می‌گیرد.

سیستم نمایه‌سازی دلبیواس درباره موضوع مجموعه مدارک، ساختار انگلیسی یا تئوری نمایه‌سازی از

### پی‌نوشت‌ها:

1. Online Computer Library Center
2. Ohio College Library Center
3. Ohio State University
4. <http://orc.rsch.oclc.org:5061/>

### فهرست منابع

- آرتانندی، سوزان. (۱۳۶۳). *نمایه‌سازی همارا*. ترجمه فیروزان زهادی. اطلاع‌رسانی نشریه فنی مرکز اسناد و مدارک علمی، دوره هفتم (ش ۲)، ص ۳۳-۴۱.
- اسونیوس، الین. (۱۳۷۲). *پیش‌همارائی آری یا نه؟* ترجمه علی مزینانی. فصلنامه کتاب. دوره چهارم، شماره چهارم، ص ۴۷-۵۴.
- حریری، مهرانگیز. توفیق، مسعوده. (۱۳۶۲). *نمایه‌گردان*. اطلاع‌رسانی. دوره هشتم (شماره ۲)، ص ۱-۳۰.
- راولی، جنیفر. (۱۳۷۴). *نمایه‌سازی و چکیده‌نویسی*. ترجمه دکتر جعفر مهرداد. تهران، سازمان مدارک فرهنگی انقلاب اسلامی.
- سینائی، علی. (۱۳۵۱). *نمایه‌سازی همارا*. نشریه فنی مرکز مدارک علمی. دوره ۱ (شماره ۲)، ص ۷-۱۴.

فاگمن، رابرت. (۱۳۷۴). *تحلیل موضوعی و نمایه‌سازی، مبانی نظری و توصیه‌های عملی*. ترجمه علی مزینانی. تهران، کتابخانه ملی جمهوری اسلامی ایران.

Bhattacharyya, K.(1974) “*The effectiveness of Natural Language In Science Indexing And Retrival*”. Journal of Documentation, vol. 30, No.3, sep. p. 235-254.

Brodie, Nanacy E.(1970). “*Evaluation of a kwic index for library literature*”. Journal of American Society for Information Science. January-February, pp.22-28.

Burress, Elain p.(1980) “Automated indexing versus kowc: A performance comparison. Journal of American Society for Information Science. pp.60-63.

Chan, Lois Mai. “Exploiting LCSH, LCC, and DDC to Retrieve Networked Resources. Available at <http://lcweb.loc.gov/vatdir/bibcontrol/chan-paper.htm>

Crook, M. Annual Report. “OCLC History and Mission The OCLC Online Union Catalog: and intcomparable library resource. Publishing Research Quarterly, 11(3) Fall 95, p.39-50. available at <http://www.oclc.org/oclc/ar95/profhist.hist.htm>

Davis, Charles H.(1978) “Evidence of OCLC’s Potential for Spetial Libraries and Technological Information Centers. JASIS. pp.255-256.

Dillon, Martin. (1982) “*Automatic Classification of Harris Survey Questions: An Experiment in the Organization of Information*”. Journal of American society for Information Science. pp. 294-301.

Dillon, Mqrtin. Macdonald, Laura K. (1983) “*Fully Automatic Book Indexing*”. Journal of documentation.vol.39,No.3. pp.135-155.

Doszkocs, Tamas E.(1986) “*Natural Language Processing in information Retrieval*”. Journal of American Society for Information Science. 37(4): pp.191-196.

Dutta, S. & P.K. (1984) “*Pragmatic Approach to Subject Indexing: A new Concept*”. Journal of American Science for Information Science. 35(6): pp.323-331.

Ford, Karin. “*Cooperative Library Services*” 2000. available at <http://www.oclc.org/western/services/index.htm>

Godby, Jean.(1998) “*WordSmith: Research Project bridges gap between tokens and indexes*.”Accessing OCLC in the 21th century. OCLC Newsletter. No.234.

Godby, C.Jean. “*The WordSmith Indexing System*”. Available at h <http://www.oclc.org/research/publications/arr/1998/godby-reighart/wordsmith.htm>

Grishman, Ralph. (1984) “*Natural Language Processing*”. Journal of American Science for Information Science. 35(5): pp. 291-296.

Hersey, D.F. et al.(1971) “*Free Text Word Retrieval And Scientist Indexing: Performance Profiles and Costs*”. Journal of Documentation. Vol.27.No.3, pp.167-183.

- Hsieh-Yee, Ingrid.(1996) "*The Cataloging Practices of Spetial Libraries and their Relationship with OCLC*". Spetial Libraries Association, vol. 87,No.1, pp.10-20.
- Hysell, D. (1996) "*CLC's website grows*". OCLC Newsletter, (222) o.10.
- Jahoda, Gerald. (1970) "*Information Storage and Retrieval Systems for individual Researchers*". NewYork: Wiley-nterscience..pp.1-111.
- Jolley, J.L.(1976) "*The Terminology of Coordinate Indexing*". Aslib Proceedings. 28(3), pp.120-128,
- Jordan, Jay. (1998). "*Helping Libraries Participate in the Global, Digital Community*". OCLC Newsletter 234,No.234.available at [www.oclc.org/oclc/new/n234/mem-users-council-hold-elections-htm](http://www.oclc.org/oclc/new/n234/mem-users-council-hold-elections-htm).
- Katz, W.A. (1974) "*Introduction to Reference Work*". New York: McGrawHill.
- Lancaster, F. Wilfrid. (1979) "*Information Retrieval Systems: Characteristics, Testing and Evaluation*". 2<sup>nd</sup> ed. NewYork: John wiley.
- Lancaster, F.W. (1991) "*Indexing and Abstracting in Theory and Practice*". London: Library Association.
- Libraries, Internet and oclc (2002). available at <http://www.oclc.org/oclc.promo/9768110/110.htm>
- OCLC announces electronic collection online Advanced Technology Libraries. 25(8) Aug.96,p.1-2.
- OCLC News Releases/OCLC Users Council. "*OCLC Users Council meets to discuss regional networks, Consortia and other partnerships*". Available at <http://www.oclc.org/oclc/press/96103.ibm.htm>
- Promenschenkel, George. Oluc 25:Home Sweet OCLC. Available at [www.oclc.org/oclc/new/n221/oluc.htm](http://www.oclc.org/oclc/new/n221/oluc.htm)
- Promenschenkel, George, Growing oclc.newsletter 224. available at [www.oclc.org/oclc/osu95/grow.htm](http://www.oclc.org/oclc/osu95/grow.htm)
- Ramsden, Michael J. (1974) "*An Introduction to Index Language Construction: A Programmed Text*". London: Clive Bingley.
- Roughton, Karen.(1982) "*Thinking of OCLC Serials Control? Read this.*"The Serials Librarian.vol.7(1). pp.23-30.
- Rowley, Jennifer. (1982). "*Abstracting and Indexing*". London: Clive bingley.
- Shafer, K. Scorpion Pfoject explorers using Dewey to organize the web. OCLC Newsletter, (222) Jul/Aug 96, p.20-21.