

نمایه‌سازی

Indexing

مهندس حمید رضا نیرومند

<http://niroomand.ir>

توجه:

- استفاده از جزوه بدون حضور در کلاس‌ها پیشنهاد نمی‌شود.
- این یک انتشار غیررسمی از جزوه و ویژه دانشجویان مهندس نیرومند است.
- این، نسخه ۱,۰,۰ از جزوه است و به مرور غنی‌تر خواهد شد. (لطفاً اشتباهات سهوی که مشاهده می‌کنید را به ایمیل info@niroomand.ir ارسال نمایید).

فهرست

۵.....	انواع شیوه‌های اطلاع‌رسانی
۵.....	نمایه‌سازی
۵.....	تعریف نمایه
۵.....	دلایل نیاز به نمایه‌سازی
۵.....	اهداف نمایه‌سازی
۶.....	تعریف فراداده یا Metadata
۷.....	معرفی برخی متا تگ‌ها در طراحی وب
۸.....	نحوه به دست آوردن کلیدواژه (کلمات کلیدی)
۸.....	روش‌های به دست آوردن کلمه کلیدی
۸.....	فراوانی وزنی یا tf-idf weight
۱۱.....	الگوریتم یافتن مقالات مرتبط با یک مقاله از دیتابیس با توجه به عنوان مقالات
۱۵.....	تاریخچه نمایه‌سازی
۱۶.....	نمایه‌سازان
۱۶.....	نمایه‌سازان چه کاری انجام می‌دهند؟
۱۶.....	نمایه‌ساز به چه مواردی توجه می‌کند؟
۱۶.....	ویژگی نمایه‌سازان
۱۷.....	زبان‌های نمایه‌سازی
۱۷.....	انواع زبان‌های نمایه‌سازی از لحاظ انتخاب مدخل
۱۸.....	انواع نمایه از لحاظ روش تنظیم
۱۹.....	انواع نمایه‌سازی از لحاظ همارایی
۲۰.....	انواع نمایه‌سازی از لحاظ روش تهیه
۲۰.....	دلایل نیاز به نمایه‌سازی خودکار
۲۰.....	نرم افزارهای نمایه‌سازی خودکار
۲۰.....	آشنایی با برخی نمایه‌های قدیمی
۲۰.....	نمایه‌های گردان
۲۰.....	۱- نمایه کوئیک (KWIC)
۲۲.....	۲- نمایه کووک (KWOC)
۲۳.....	۳- نمایه جایگشتی (Permuterm)
۲۵.....	نقاط قوت و ضعف نمایه‌ی گردان
۲۶.....	انواع نمایه انتهایی کتاب
۲۷.....	تفاوت نمایه انتهایی کتاب، درون بافتی و برون بافتی
۲۷.....	اجزای نمایه‌های درون بافتی و برون بافتی
۲۸.....	اصطلاح مرجح چیست؟
۲۸.....	ملاک‌های انتخاب اصطلاح مرجح
۲۸.....	گام‌های ایجاد یک نمایه به روش دستی
۲۹.....	نکاتی در مورد ساختار زبان شناسی مدخل
۲۹.....	روال تهیه یا خط مشی نمایه

۳۰ معیارهای ارزیابی نمایه
۳۰ از نظر آقای کلوند (Cleveland) در ارزیابی نمایه‌ها چه نکاتی باید مد نظر قرار گیرد؟
۳۱ ویژگی‌های یک نمایه خوب چیست؟
۳۱ عوامل مؤثر در نمایه‌سازی
۳۲ گنج‌واژه یا اصطلاح‌نامه
۳۲ اهداف اصطلاح‌نامه
۳۲ روابط میان اصطلاحات
۳۳ انواع اصطلاح
۳۴ انواع رابطه سلسه مراتبی
۳۵ تفاوت واژه‌نامه و اصطلاح‌نامه
۳۶ چکیده یا Abstract
۳۶ انواع چکیده بر حسب لحن
۳۷ انواع چکیده بر حسب حجم
۳۸ انواع چکیده بر حسب تهیه‌کننده
۳۹ نکات مهم در تهیه چکیده مقالات و پایان‌نامه‌ها و طرح‌های پژوهشی
۳۹ گام‌های تهیه چکیده
۴۰ رده‌بندی چیست؟
۴۰ انواع رده‌بندی
۴۳ روش‌های کمک گرفتن از یک کتابدار (مفهوم Ask a Librarian)
۴۵ آنتولوژی چیست؟
۴۵ انواع وب
۴۵ ۱- وب ۱ (وب سنتی) (Traditional Web)
۴۵ ۲- وب ۲ (وب اشتراکی) (Social Web)
۴۵ ۳- وب ۳ (وب معنایی یا وب هوشمند) (Semantic Web)
۴۵ روش‌های معنا بخشیدن به صفحات وب
۴۶ XML چیست؟
۴۸ URI چیست؟
۴۹ RDF چیست؟
۴۹ مزایای RDF نسبت به XML
۵۰ روش اعمال RDF به یک صفحه با HTML5
۵۱ مدل مخفی مارکوف چیست؟
۵۳ کاربردهای مدل مخفی مارکوف
۵۵ نکاتی در مورد موتورهای جستجو
۵۵ ۳ نکته مهم در ساخت یک موتور جستجوی عالی
۵۶ معرفی چند سایت پژوهشی ایرانی و خارجی
۵۶ سایت‌های خارجی
۵۷ سایت‌های ایرانی

انواع شیوه‌های اطلاع‌رسانی

- ۱- نمایه‌سازی
- ۲- چکیده‌نویسی
- ۳- رده‌بندی

نمایه‌سازی

تعریف نمایه^۱

An Index is a list of words or phrases (heading) and associated Pointers (locators) to where useful material relating to that heading can be found in a Document.

نمایه، لیستی است از لغات یا عبارات (که به آن سرآیند یا heading گفته می‌شود) به انضمام اشاره‌گرها (که به آن locators یا «جای‌نما» گفته می‌شود) به محل‌هایی که مطالب مفید و مرتبط با آن سرآیند در سند مورد نظر قابل مشاهده است.

تعاریف دیگر از نمایه

- به طور ساده نمایه، سیاهه یا لیستی نظام یافته، برای هدایت منظم یک متن، محتوا، مجموعه‌ای از مدارک یا هر گونه اطلاعات ضبط شده است.
- نمایه استفاده‌کنندگان را از اطلاعات معلوم به اطلاعات اضافی ناشناخته هدایت می‌کند.

دلایل نیاز به نمایه‌سازی

- ۱- تحلیل مؤثر و سریع اسناد
- ۲- افزایش حجم اطلاعات
- ۳- سرعت پایین بازیابی اطلاعات در روش‌های کتابداری سنتی
- ۴- تغییر روش‌های ذخیره و بازیابی اطلاعات

اهداف نمایه‌سازی

- ۱- زمان و تلاش برای یافتن اطلاعات را به حداقل میزان کاهش دهد.
- ۲- موفقیت جستجوی استفاده‌کننده را به حداکثر برساند.

^۱ Index

- ۳- برقراری ارتباط میان مفاهیم
- ۴- تنظیم شناسه‌ها به ترتیبی نظام‌مند و مؤثر
- ۵- شناسایی سریع مدارک در یک مجموعه
- ۶- سازماندهی اطلاعات به قصد بازیابی سریع و آسان مدارک

به طور خلاصه، هدف نمایه‌سازی عبارت است از: انتخاب واژه مناسب با زمان کم و تلاش کمتر

نمایه یک فرا داده است

تعریف فراداده یا Metadata

به اطلاعات اضافه که درباره یک داده ثبت می‌شود تا در طبقه‌بندی از آن‌ها استفاده شود گفته می‌شود. (داده‌ای در مورد داده)

مثال ۱: مشخصات عکسی که توسط دوربین دیجیتال گرفته می‌شود (مانند زمان گرفتن عکس، سایز عکس، نام و مدل دوربین، نام و مدل لنز دوربین، موقعیت جغرافیایی محل گرفتن عکس و ...) در header عکس ذخیره می‌شود. این اطلاعات به دسته‌بندی عکس توسط برنامه‌های سیستم عامل کمک می‌کند. (به طور مثال در سیستم عامل مک، عکس‌ها به طور خودکار نسبت به محلی که گرفته شده‌اند دسته‌بندی می‌شوند: عکس‌های شمال، عکس‌های مشهد و...)

مثال ۲: در طراحی سایت، اطلاعاتی اضافه در مورد هر صفحه در بین تگ <head> در داخل یک <meta> قرار داده می‌شود. این اطلاعات به موتورهای جستجو در دسته‌بندی سایت‌ها کمک می‌کند.

معرفی برخی متا تگ‌ها در طراحی وب

```
1 <!DOCTYPE html>
2 <html>
3 <head>
4 <title>Aftabgardan </title>
5 <meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
6 <meta http-equiv="expires" content="0" />
7 <meta name="resource-type" content="document" />
8 <meta name="distribution" content="global" />
9 <meta name="author" content="Aftabgardan" />
10 <meta name="copyright" content="Copyright (c) by Aftabgardan" />
11 <meta name="keywords" content="انگلیسی, مستند, موبایل, آموزش, دانلود, تفریح, نرم افزار" />
12 <meta name="description" content="این فرهنگی مجازی با هدف فراهم کردن محیطی لذتبخش و همچنین در دسترس قرار دادن آنچه یک کاربر رایانه نیاز دارد، اعم از دانلود نرم افزار ... ایجاد شده است." />
13 <meta name="robots" content="INDEX, FOLLOW" />
14 <meta name="revisit-after" content="1 DAYS" />
15 <meta name="rating" content="general" />
16 <meta name="generator" content="PHP-Nuke" />
17 </head>
```

1- <meta name="description" content="">

این متا تگ حاوی توضیح کوتاهی درباره هدف سایت است که معمولاً در نتایج جستجو در زیر عنوان هر نتیجه نمایش داده می‌شود.

2- <meta name="keywords" content="keyword1,keyword2, ...">

حاوی نهایتاً ده کلمه کلیدی درباره هر صفحه‌ی وب. کلماتی که بیشترین ارتباط را با موضوع مطالب صفحه دارند و با کاما هم جدا شده‌اند.

3- <meta name="author" content="Hamid Reza Niroomand">

حاوی نام نویسنده مقاله موجود در صفحه جاری

4- <meta name="rating" content="general">

حاوی رده‌ی سنی مرتبط با محتوای صفحه.

5- <meta name="generator" content="PHP-Nuke">

نام CMS یا سیستم مدیریت که توسط آن، سایت ایجاد شده است.

6- <meta charset="UTF-8">

حاوی مجموعه کارکتهایی که در صفحه به کار رفته است. (معمولاً utf-8)

7- <meta name="distribution" content="global">

حاوی منطقه جغرافیایی مرتبط با محتوای صفحه

8- <meta http-equiv="expires" content="0">

حاوی تاریخ انقضای مطالب صفحه

نحوه به دست آوردن کلیدواژه (کلمات کلیدی)

شاید بتوان گفت مهم‌ترین موضوع در نمایه‌سازی، انتخاب «کلیدواژه» به عنوان مدخل است چرا که همه واژه‌های یک مقاله یا عنوان آن مقاله از نظر نویسنده و جوینده، نقش کلیدی ندارند و در بعضی از آن‌ها کلماتی مانند حروف اضافه (مانند «و»، «به»، «از» و ...) و حروف ربط (مانند «بنابراین»، «که»، «پس» و ...) و ضمائر (مانند «من»، «او» و ...) و حروف تعریف (مانند «عبارت است از»، «یعنی» و ...) و نظایر آن وجود دارد که طبعاً حامل هیچ گونه اطلاعاتی در مورد موضوع و محتوای سند نمی‌باشد و هیچ وقت مورد جستجو واقع نمی‌شوند. بنابراین این نوع کلمات نباید به عنوان مدخل در نمایه‌نامه قرار گیرند.

معمولاً لیستی از واژه‌های غیر مجاز یا بازدارنده (stop word یا stop list) تهیه می‌شود تا با بازداشتن کامپیوتر یا نمایه‌ساز از انتخاب این واژه‌ها:

- اولاً از انتخاب مدخل‌های بی‌فایده و در نتیجه حجیم شدن بی‌دلیل نمایه‌نامه و افزایش قیمت آن جلوگیری شود.
 - ثانیاً با کاهش مدخل‌های بی‌ربط از بار زحمتی که در جستجوی موضوعات بر دوش جستجوگر گذاشته می‌شود تا حد زیادی کاسته شود.
- این لیست‌ها معمولاً با توجه به ماهیت و موضوع رشته‌ی مورد نظر فراهم می‌شود.

روش‌های به دست آوردن کلمه کلیدی

۱- روش دستی (انسانی):

در این روش، کلمات کلیدی توسط نمایه‌ساز (یعنی یک انسان) و طبق تجربه‌ی شخصی او انتخاب می‌شوند.

۲- روش ماشینی (کامپیوتری):

در این روش، کلمات کلیدی توسط کامپیوتر و با استفاده از یک الگوریتم به دست می‌آیند.

یکی از روش‌های به دست آوردن کلمات کلیدی و غیرکلیدی، استفاده از الگوریتمی با نام «فراوانی وزنی» است.

فراوانی وزنی یا tf-idf weight

مخفف term frequency-inverse document frequency و به معنی «فراوانی کلمه، برعکس فراوانی سند» و الگوریتمی برای پیدا کردن کلمات کلیدی یک سری مقاله با استفاده از متن خود مقالات است.

این سیستم وزنی نشان می‌دهد چقدر یک کلمه برای یک سند مهم است. این کار را از طریق وزن دادن به لغات بر اساس فراوانی آن در سند انجام می‌دهد (یعنی هر چه فراوانی یک لغت در یک سند بیشتر باشد وزن آن بیشتر است).

idf: معیاری است برای به دست آوردن فراوانی یک کلمه نسبت به کل مقالات:

$$\text{idf}(t, D) = \log \frac{N}{|\{d \in D : t \in d\}|}$$

این فرمول یعنی idf مربوط به یک کلمه، به این صورت محاسبه می‌شود:

لگاریتم [تعداد کل مقالات، تقسیم بر (تعداد مقالاتی که شامل این کلمه می‌شوند+1)]

این معیار به تنهایی برای یافتن کلمات بازدارنده کافی است! فرض کنید کل مقالات شما ۱۰۰ تا باشد و کلمه‌ی «در» در همه مقالات تکرار شده باشد. در این صورت لگاریتم ۱۰۰ تقسیم بر ۱۰۰ (یعنی لگاریتم ۱) می‌شود: صفر. هر گاه این معیار، صفر یا خیلی نزدیک به صفر شد یعنی آن کلمه جزء کلمات بازدارنده است.

tf: برای به دست آوردن tf مربوط به یک کلمه در یک دیتابیس، باید ببینید کدام سندها حاوی آن کلمه هستند و در هر سند آن کلمه چند بار تکرار شده است؟ ماکزیمم تعدادی که این کلمه در یک سند تکرار شده را tf بگیرید. به مثال زیر دقت کنید:

Document 1		Document 2	
Term	Term Count	Term	Term Count
this	1	this	1
is	1	is	1
a	2	another	2
sample	1	example	3

در این مثال دو سند داریم که تعداد تکرار کلمات آن سند مقابل هر کلمه نوشته شده. اگر شما tf مربوط به کلمه example را بخواهید، پاسخ، ۳ خواهد بود. چون این کلمه در سند دوم بیشترین تکرار را داشته که ۳ بار بوده. و یا tf کلمه a برابر با ۲ می‌شود چون ۲ بار در سند اول تکرار شده.

tfidf: این معیار که پارامتر تصمیم‌گیری در مورد کلیدی بودن یا نبودن یک کلمه است از طریق ضرب دو پارامتر بالا به دست می‌آید:

$$\text{tfidf}(t, d, D) = \text{tf}(t, d) \times \text{idf}(t, D)$$

شما با کنترل **tfidf** می‌توانید انواع کلمه (کلیدی، غیر کلیدی و بازدارنده) را به دست آورید.

مثال:

فرض کنید ۴ مقاله‌ی زیر در دیتابیس موجود باشد و کاربری عبارت «آثار رکود اقتصادی در رضایت‌مندی مردم ایران» را جستجو کند. با استفاده از روش فراوانی وزنی مرتبط‌ترین مقاله با این عبارت را بیابید.

۱۰۱ اقتصاد مسکن در ایران

۱۰۲ بررسی آثار اقتصادی رکود بازار فرش

۱۰۳ نقش فرش در اقتصاد ایران

۱۰۴ عوامل اقتصادی مهاجرت از روستاها به شهرها

فرمول:

$$\text{فراوانی وزنی} = \log \frac{\text{تعداد کل مقالات}}{\text{تعداد مقالاتی که کلمه مورد نظر در آن‌ها تکرار شده} + 1}$$

به دست آوردن idf کلمات:

$$\text{idf}_{\text{آثار}} = \log \frac{4}{1+1} = 0.3$$

$$\text{idf}_{\text{رکود}} = \log \frac{4}{1+1} = 0.3$$

$$\text{idf}_{\text{اقتصادی}} = \log \frac{4}{2+1} = 0.12$$

$$\text{idf}_{\text{در}} = \log \frac{4}{2+1} = 0.12$$

$$\text{idf}_{\text{رضایت‌مندی}} = \log \frac{4}{0+1} = 0.60$$

$$\text{idf}_{\text{مردم}} = \log \frac{4}{0+1} = 0.60$$

$$\text{idf}_{\text{ایران}} = \log \frac{4}{2+1} = 0.12$$

پیدا کردن tf کلمات:

$$\text{tf}_{\text{آثار}} = 1$$

$$\text{tf}_{\text{رکود}} = 1$$

$$tf_{\text{اقتصادی}} = 1$$

$$tf_{\text{در}} = 1$$

$$tf_{\text{رضایت‌مندی}} = 0$$

$$tf_{\text{مردم}} = 0$$

$$tf_{\text{ایران}} = 1$$

به دست آوردن **tfidf** کلمات:

$$tfidf_{\text{اثر}} = 1 * 0.3$$

$$tfidf_{\text{رکود}} = 1 * 0.3$$

$$tfidf_{\text{اقتصادی}} = 1 * 0.12$$

$$tfidf_{\text{در}} = 1 * 0.12$$

$$tfidf_{\text{رضایت‌مندی}} = 0 * 0.60$$

$$tfidf_{\text{مردم}} = 0 * 0.60$$

$$tfidf_{\text{ایران}} = 1 * 0.3$$

نتیجه: کلماتی که **tfidf** بیشتری دارند، کلیدی‌تر به حساب می‌آیند. بنابراین، کلمات «آثار» و «رکود» کلمات کلیدی مورد جستجوی کاربر هستند و مقاله ۱۰۲ مرتبط‌ترین مقاله با جستجوی کاربر است. (چون هر دوی این کلمات را دارد)

الگوریتم یافتن مقالات مرتبط با یک مقاله از دیتابیس با توجه به عنوان مقالات

- **گام اول:** ابتدا باید عنوان مطلب فعلی را با معیار فاصله (Space) تکه‌تکه کنید. مثلاً فرض کنید مطلبی دارید با عنوان «آموزش ترمیم عکس قدیمی در فتوشاپ». باید با توابعی که در اختیار دارید (مانند تابع `explode` در زبان PHP و `Split` در زبان‌های دات‌نت) عنوان را با معیار فاصله در اصطلاح منفجر کنید تا هر کلمه در یک خانه از یک آرایه قرار گیرد. در انتهای این مرحله یک آرایه خواهید داشت که در هر خانه یک کلمه از کلمات عنوان مطلب وجود دارد.
- **گام دوم:** باید کلمات غیرکلیدی از آرایه حذف شوند و نهایتاً فقط کلمات کلیدی باقی بمانند. برای اینکه این کار دقیق‌تر صورت گیرد شما به سه مجموعه داده (دیتابیس) نیاز دارید:
- دیتابسی به نام **Stop Words** یا **کلمات بازدارنده:**

به حروف اضافه (مانند «به»، «از» و ...) و حروف ربط (مانند «که»، «پس» و ...) و ضمائر (مانند «من»، «او» و ...) و همینطور افعال عمومی مثل «باش، کن و ...» که در تمامی متون یافت می‌شوند در اصطلاح کلمات بازدارنده یا غیرمجاز یا Stop List یا Stop Words گفته می‌شود.

- دیتابیس‌ی به نام **Go List** یا **Keywords** یا **کلمات کلیدی**:

این دیتابیس حاوی کلماتی خواهد بود که در یک زمینه خاص، کلمه کلیدی هستند. مثلاً اگر مطالب یک سایت، در زمینه رشته‌ی کامپیوتر است، کلمات کلیدی آن با یک سایت خبری سیاسی متفاوت است. شما باید هر کلمه‌ای که در این دیتابیس نیست را هم حذف کنید.

به طور مثال در مطلب «آموزش ترمیم عکس قدیمی در فتوشاپ» کلماتی مثل «ترمیم» و «قدیمی» کلمه کلیدی نیستند و باید با توجه به دیتابیس کلمات کلیدی رشته کامپیوتر حذف شوند.

- دیتابیس‌ی به نام **Semi-Stop Words** یا **کلمات نیمه‌بازدارنده**:

- کلماتی که جزء دیتابیس کلمات بازدارنده و کلمات کلیدی نباشند، کلمات نیمه‌بازدارنده یا نیمه مجاز به حساب می‌آیند.

کلمات نیمه‌بازدارنده نیز باید از آرایه حذف شوند. مانند «ترمیم» و «قدیمی» در عنوان فرضی ما. در حقیقت شما به یکی از دیتابیس‌های **Go List** و **Semi-Stop Words** نیاز دارید و دومی می‌تواند از روی دو تای دیگر به دست آید. یعنی بعد از اینکه شما کلمات بازدارنده را حذف کردید و از روی دیتابیس کلمات کلیدی، کلمات کلیدی را مشخص کردید، هر چه باقی می‌ماند می‌شود کلمات نیمه‌بازدارنده که باید آن‌ها نیز حذف شوند.

- **گام سوم:** در انتهای گام ۲ شما یک آرایه دارید که در آن کلمات کلیدی که در عنوان مطلب فعلی بوده، قرار دارد. مثلاً در عنوان فرضی ما کلمه «در» با استفاده از **Stop Words** حذف می‌شود، کلمات «ترمیم» و «قدیمی» هم با توجه به دیتابیس **Semi-Stop Words** حذف می‌شوند و نهایتاً کلمات «آموزش» و «عکس» و «فتوشاپ» در آرایه می‌ماند.

- حالا باید این کلمات را با استفاده از عملگر **LIKE**، یک به یک در دیتابیس مقالات، جستجو کنید. (مثلاً یک بار «آموزش» را... یک بار «عکس» را...) نتیجه این جستجو لیستی از مقالاتی خواهد بود که مرتبط با هر کلمه هستند. مثلاً چیزی شبیه به تصویر زیر خواهیم داشت:

فتوشاپ	عکس	آموزش	شماره مقالات
1	1	2	
2	2	5	
8	7	7	
10	8	8	
		10	

- حالا فرض کنید قرار است نهایتاً ۵ مقاله مرتبط را لیست کنید. این سؤال پیش می‌آید که کدام مقالات را بالاتر بیاوریم. (دقیقاً همان سؤالی که گوگل هنگام یافتن هزاران نتیجه در قبال جستجوی شما با آن مواجه است و باید فکری در موردش بکند. یعنی کدام سایت را بالاتر بیاورد؟)
- باید ابتدا مقالاتی آورده شوند که تعداد بیشتری از کلمات کلیدی‌ای که در مرحله قبل به دست آمد را شامل می‌شوند. به طور مثال در تصویر بالا، مقالات شماره ۲ و ۸ هر سه کلمه را شامل می‌شوند، پس ابتدای لیست می‌آیند. سپس از راست به چپ (و اگر زبان جستجو، انگلیسی بود، از چپ به راست) اولویت کلمات را بالاتر فرض کنید و هر مقاله‌ای که تعداد بیشتری کلمه کلیدی را دارد لیست کنید. مثلاً مقاله شماره ۷ که کلمات «آموزش» و «عکس» را دارد در جایگاه بعدی قرار می‌گیرد و مقاله شماره ۱۰ در جایگاه بعد و در نهایت مقاله شماره ۵ که فقط یک کلمه کلیدی را دارد در لیست قرار می‌گیرد.
- بنابراین ترتیب مقالات مرتبط به صورت زیر خواهد بود:
 - مقاله شماره ۲
 - مقاله شماره ۸
 - مقاله شماره ۷
 - مقاله شماره ۱۰
 - مقاله شماره ۵

شکل ساده‌ای از این پیاده‌سازی به زبان PHP در ادامه آمده است:

```

//////////Find Related Articles
$title_words = explode(' ', $title); ← انفجار عنوان مقاله فعلی
$stopwords = $db->sql_fetchrow($db->sql_query("select `stopword` from `stopwords`"));
$related_articles = "";
foreach ($title_words as $title_word) { ← حذف کلمات بازدارنده
    if(!in_array($title_word, $stopwords))
    {
        ← یافتن مقالاتی که حاوی هر کلمه هستند
        $result=mysql_fetch_row($db->sql_query("select `sid` from `".$prefix."_stories` WHERE `title` LIKE '%$title_word%'"));
        $related_articles.=(!empty($related_articles)?','.'').implode(',', $result);
    }
}
$related_articles=explode(',', $related_articles);
$related_articles=array_filter($related_articles, 'strlen');
$related_articles = array_count_values($related_articles); ← شمردن تعداد کلمات کلیدی در عنوان مقالات یافت شده
arsort($related_articles); ← مرتب‌سازی مقالات یافت شده بر اساس تعداد کلمات کلیدی موجود در آنها
$i=1;
echo '
    <ul>';
foreach ($related_articles as $key => $value) { ← نمایش ۵ مقاله اول لیست
    if($key!=$sid)
    {
        $related_article = $db->sql_fetchrow($db->sql_query("select `title` from `".$prefix."_stories` WHERE `sid`=$key"));
        echo '
            <li><a href="/article/'.$key.'">'.$related_article['title'].'</a></li>
        ';
    }
    if ($i==5)
        break;
    $i++;
}
echo '
    </ul>
</div>';
//////////

```

تاریخچه نمایه‌سازی

- تاریخ نمایه‌سازی و چکیده‌نویسی با تاریخ نوشتن ارتباط نزدیکی دارد.
- از اولین نمایه‌ها می‌توان به کشف اللغات (concordance) اشاره کرد.
- اختراع چاپ، رشد سریع چاپ و تولید کتاب، نیاز به نمایه‌های کتاب را افزایش داد.
- در قرن ۱۶ نمایه‌های با کیفیت برای کتاب ایجاد شد.
- در قرن ۱۷ نوع جدید ابزار اطلاعاتی (نشریات ادواری) به وجود آمد و با رشد سریع آن، نمایه‌ها ضرورت یافتند.
- همزمان با رشد مجلات علمی مجلات چکیده نیز توسعه یافت.
- دهه ۱۹۴۰ دهه انفجار اطلاعات بود و بیش از گذشته نیاز به نمایه‌ها احساس شد.
- کتاب «المعجم المُفهرَس» نمایه آیات قرآن است.

نمایه‌سازان

نمایه‌سازان چه کاری انجام می‌دهند؟

تحلیل سند و تخصیص واژه‌های موضوعی که فکر می‌کنند استفاده‌کننده بر اساس آن واژه‌ها، جستجو خواهد کرد.

نمایه‌ساز به چه مواردی توجه می‌کند؟

۱- توجه به اصول و خط مشی نمایه‌سازی

۲- نیاز استفاده‌کنندگان از نمایه

ویژگی نمایه‌سازان

۱- نظم و انضباط

۲- علاقه به جزئیات

۳- حافظه خوب (که یک سرمایه است!) (در صورت نیاز، از نرم افزارهایی که به جای حافظه قابل استفاده

است مانند excel و access استفاده شود)

۴- سریع خواندن (که یک ویژگی مثبت است)

۵- نمایه‌سازان سانسورکننده خوبی نیز هستند (چه بسا به طور غیر عمد) (گاهی باید برخی بخش‌های یک

سند را نادیده گرفت)

۶- ذهن کنجکاو و خلاق دارند؛ به طور مثال از خود می‌پرسند: اگر این اصطلاح را انتخاب کنم، آیا کاربران

در جستجو دقیقاً همین واژه را استفاده می‌کنند؟

۷- نمایه‌ساز خوب، خود را به جای مخاطب قرار می‌دهد.

زبان‌های نمایه‌سازی

- زبان نمایه‌سازی زبانی است که برای توصیف موضوع یا سایر جنبه‌های اطلاعات یا مدارک در نمایه به کار می‌رود.
- زبان نمایه‌سازی به زبان ساختگی و قراردادی اطلاق می‌شود که برای مقاصد نمایه‌سازی به ویژه قابلیت بازیابی اطلاعات و مدارک به کار گرفته می‌شود.

انواع زبان‌های نمایه‌سازی از لحاظ انتخاب مدخل

۱- زبان آزاد:

در این زبان نمایه‌ساز آزاد است از هر واژه‌ای که بتواند اطلاعات مدرک را برساند برای انتخاب به عنوان مدخل کمک گیرد. خواه این واژه‌ها در مدرک باشد یا نباشد.

مثال: اگر یک سند حاوی جمله «امروزه علم کامپیوتر یکی از کاربردی‌ترین علوم به حساب می‌آید» باشد، کلماتی مانند: علم (که در سند موجود است)، دانش (که در سند موجود نیست)، کامپیوتر (که در سند موجود است)، رایانه (که در سند موجود نیست) و... می‌توانند مدخل قرار گیرند.

۲- زبان طبیعی:

اگر کلماتی که نمایه‌ساز انتخاب کند در سند موجود باشد این نمایه‌سازی به زبان طبیعی مشهور است.

مثال: اگر یک سند حاوی جمله «امروزه علم کامپیوتر یکی از کاربردی‌ترین علوم به حساب می‌آید» باشد، فقط کلماتی مانند: علم، کامپیوتر، کاربردی (که در سند موجود است) می‌توانند مدخل قرار گیرند.

۳- زبان مقید یا مهارشده یا کنترل‌شده:

در این نمایه، نمایه‌ساز بعد از درک مدارک و موضوع باید از بین واژه‌ها و اصطلاحاتی که از قبل تهیه و اعتبارسنجی و وضع شده مناسب‌ترین واژه را برای موضوع مدرک انتخاب کند. بنابراین نمایه‌ساز فقط قدرت انتخاب دارد و نه بیش‌تر و دیگر آن آزادی عمل را در گزینش واژه ندارد.

مثال: اگر یک سند حاوی جمله «امروزه علم کامپیوتر یکی از کاربردی‌ترین علوم به حساب می‌آید» باشد و نمایه‌ساز مقید شده باشد که فقط از کلمات کاملاً فارسی کمک بگیرد، فقط کلماتی مانند: دانش، رایانه، کاربردی (که کاملاً فارسی هستند) می‌توانند مدخل قرار گیرند.

مزایای زبان طبیعی

- تنوع واژه‌ها: چون نمایه‌ساز آزاد است از هر واژه‌ای استفاده کند.
- عدم مواجه شدن با کمبود اصطلاح.

- هم‌خوان بودن زبان نمایه‌ساز با زبان مدرک: از آن‌جا که نمایه‌ساز آزاد است هر کلمه‌ای را انتخاب کند مجبور نیست اصطلاحاتی را انتخاب کند که با موضوع مدرک نزدیک نیست و هم‌سویی ندارد.

معایب زبان طبیعی

- از هم‌گسیختگی اصطلاحات و نمایه‌ها که دقیقاً نقطه مقابل نظام زبان مهار شده است.
- به دلیل تنوع اصطلاحات ما با شماری از نمایه‌های تک مدخلی برای ورود به اطلاعات مواجه‌ایم. در حالی که اگر از زبان کنترل شده استفاده شده باشد واژه‌های مختلفی که در مدرک وجود دارد اما همگی اشاره به یک مفهوم می‌کند، یک نمایه در نظر گرفته می‌شود بنابراین مدخل‌های چند لغت در مقابل همین یک لغت نوشته می‌شود و طبیعتاً این موضوع بازایی اطلاعات را سریع‌تر و بهتر می‌کند.
- تکثر زبان نمایه‌سازی: از آنجا که برای زبان طبیعی، استاندارد وجود ندارد، بنابراین ما با عدم وحدت میان زبان نمایه‌ساز و زبان جستجوگر مواجه هستیم که این امر سرعت و دقت در امر بازایی را کاهش می‌دهد.

مزایای زبان مقید

- نظام‌یافتگی و وحدت میان اصطلاحات (یعنی به طور مثال همیشه از کلمات فارسی به جای معادل خارجی استفاده شده) و در نتیجه بازایی کاراتر اطلاعات.
 - اشاره به مدخل‌های بیشتر در مقابل هر نمایه.
 - بازایی سریع‌تر اطلاعات به خاطر وحدت میان زبان نمایه‌ساز و جستجوگر.
- نکته: یکی از کاربردهای این مبحث در طراحی سایت‌های حاوی مطالب آموزشی است در این نوع سایت‌ها باید قبل از درج محتوا در دیتابیس و هنگام جستجو در دیتابیس حروف مشکل‌زا مانند «ی» و «ک» که (در کیبورد فارسی و عربی متفاوت هستند) و همین‌طور کلماتی مانند خرید و فروش که از نگاه مدیر سایت و از نگاه کاربر هر دو به کار گرفته می‌شوند، یکسان‌سازی یا معادل‌سازی شوند.

انواع نمایه از لحاظ روش تنظیم

- ۱- **نمایه‌های الفبایی:** مدخل‌ها به ترتیب حروف الفبا چیده می‌شوند و اکثر نمایه‌های کنونی به این صورت است.
- ۲- **نمایه‌های زمانی:** مدخل‌ها به ترتیب زمان از قدیم به حال مرتب شده و بیشتر برای کتب تاریخی به کار می‌رود.
- ۳- **رده‌ای یا موضوعی:** مدخل‌ها بر اساس عنوان موضوعات مرتب می‌شوند. بیشتر برای متون علمی به کار برده می‌شود.

نکته: می‌توانند به صورت الفبایی نیز چیده شوند که در این صورت نمایه‌ی الفبایی-رده‌ای نامیده

می‌شود. مثال: سایت noormags.ir

۴- **نمایه تکاملی:** مدخل‌ها به ترتیب تکامل چیده شده‌اند و بیشتر برای متون و کتب زمین‌شناسی مناسب است.

۵- **نمایه استنادی:** این نوع نمایه شامل یک فهرست فرعی تحت هر یک از مقالات منتشر شده است که حاوی مقالاتی است که به آن مقالات استناد کرده‌اند. به عبارت دیگر در مورد یک مقاله‌ی خاص، نمایه استنادی مشخص می‌کند که آن مقاله توسط چه مقالات دیگری که بعد از آن نوشته شده مورد استفاده قرار گرفته‌اند.

- در حقیقت بیان‌کننده ارتباط موضوعی مقاله فعلی به مقالاتی است که به آن استناد کرده‌اند.
- مزیت اصلی آن ارجاع خواننده یا استفاده‌کننده به جدیدترین مقالات است.

انواع نمایه‌سازی از لحاظ همارایی

همارایی، به ترکیب مدخل‌های نمایه گفته می‌شود.

۱- نمایه‌سازی پیش-همارا یا Pre Coordination Indexing

در این روش ترکیب عناصر تشکیل‌دهنده‌ی موضوع مورد جستجو، در هنگام نمایه‌سازی و پیش از بازیابی صورت می‌گیرد. مثال: مقالاتی در مورد کتاب‌شناسی تاریخ اصفهان به این صورت نمایه می‌شوند:

اصفهان

تاریخ

کتاب‌شناسی

مقالات...

۲- نمایه‌سازی پس-همارا یا Post Coordination Indexing

در این روش، ترکیب یا همارایی عناصر تشکیل‌دهنده‌ی موضوع مورد جستجو، در هنگام جستجو و توسط جستجوگر انجام می‌شود.

مثال: مقالاتی در مورد کتاب‌شناسی تاریخ اصفهان در مدخل‌های مختلف قرار گرفته‌اند و این به عهده جستجوگر است که بفهمد کدام مقالات در مورد موضوعی است که می‌خواهد:

اصفهان

مقالات...

تاریخ

مقالات...

کتاب‌شناسی

مقالات...

انواع نمایه‌سازی از لحاظ روش تهیه

۳- نمایه‌سازی دستی:

در این نوع نمایه‌سازی، مدخل‌ها و حتی ارجاعات توسط انسان و به صورت دستی انتخاب می‌شوند.

۴- نمایه‌سازی خودکار

فرایند استخراج مجموعه‌ای از مدخل‌های نمایه‌ای که بیانگر موضوع متن هستند توسط رایانه را نمایه‌سازی خودکار می‌گویند.

دلایل نیاز به نمایه‌سازی خودکار

- افزایش حجم اطلاعات موجود
 - تنوع ساختار در قالب‌هایی مانند مقاله، پایان‌نامه، رساله و..
 - نیاز به خلاصه‌سازی مقالات چکیده و ضرورت یکسان‌سازی قالب آن‌ها
 - ارائه درخواست‌های اطلاعاتی بهره‌گیران به صورت اصطلاح یا واژه‌های منفرد
- این موارد باعث سخت شدن استفاده از شیوه‌های دستی شده است.

نرم افزارهای نمایه‌سازی خودکار

- نرم افزار HTML/Prep
- نرم افزار HTML indexer
- نرم افزار RoboHelp HTML Edition

آشنایی با برخی نمایه‌های قدیمی

نمایه‌های گردان

نمایه‌های گردان، از قدیمی‌ترین نمایه‌ها هستند که اساس کار آن‌ها بر گردش کلید واژه‌های عنوان است. مراد از نمایه‌ی گردان نمایه‌ای است که در آن واژه‌های مهم عنوان‌های مجموعه‌ای از اسناد در نظم ویژه‌ای به گردش در می‌آید و هر یک از این واژه‌ها به ترتیب مدخلی برای ردیابی و بازیابی اطلاعات قرار می‌گیرد. بنابراین تعداد مدخل‌هایی که برای بازیابی یک سند وجود دارد حداقل به تعداد کلید واژه‌های عنوان آن است.

۱- نمایه کوئیک (KWIC)

- مخفف Key Word In Context و به معنی «کلمه کلیدی، در متن» است.

- نوعی نمایه‌ی گردان است که هر بار یکی از کلمات عنوان (به جز کلمات بازدارنده) به صورت چرخشی به عنوان مدخل قرار می‌گیرند.

- به آن در اصطلاح نمایه درون‌بافتی نیز می‌گویند چون مدخل، جزئی از بافت کلی عنوان است.

- یکی از ابتدایی‌ترین و در عین حال موفق‌ترین نمایه‌های ماشینی عنوان است.

مثال: فرض کنید ۴ مقاله با عناوین زیر داشته باشیم که شماره بازبایی آن‌ها به ترتیب ۱۰۱، ۱۰۲، ۱۰۳، ۱۰۴ است با روش کوئیک یک نمایه‌نامه برای آن‌ها تهیه کنید.

۱۰۱ اقتصاد مسکن در ایران

۱۰۲ بررسی آثار اقتصادی رکود بازار فرش

۱۰۳ نقش فرش در اقتصاد ایران

۱۰۴ عوامل اقتصادی مهاجرت از روستاها به شهرها

نمایه‌نامه:

ارجاع	بافت		
	بخش اول	مدخل	بخش دوم
۱۰۳	نقش فرش در	اقتصاد	ایران
۱۰۱		اقتصاد	مسکن در ایران
۱۰۲	بررسی آثار	اقتصادی	رکود بازار فرش
۱۰۴	عوامل	اقتصادی	مهاجرت از روستاها به شهرها
۱۰۱	اقتصاد مسکن در	ایران	
۱۰۳	نقش فرش در اقتصاد	ایران	
۱۰۲	بررسی آثار اقتصادی رکود	بازار	فرش
۱۰۲	بررسی آثار اقتصادی	رکود	بازار فرش
۱۰۴	عوامل اقتصادی مهاجرت از	روستاها	به شهرها
۱۰۴	عوامل اقتصادی مهاجرت از روستاها به	شهرها	
۱۰۲	بررسی آثار اقتصادی رکود بازار	فرش	
۱۰۳	نقش	فرش	در اقتصاد ایران
۱۰۱	اقتصاد	مسکن	در ایران
۱۰۴	عوامل اقتصادی	مهاجرت	از روستاها به شهرها
۱۰۳		نقش	فرش در اقتصاد ایران

در نمایه فوق کلمات: آثار، از، بررسی، به، در، عوامل که بار اطلاعاتی چندانی ندارند به عنوان کلمات غیر مجاز یا بازدارنده محسوب شده و کلید واژه منظور نگردیده‌اند.

۲- نمایه کووک (KWOC)

- مخفف Key Word Out of Context و به معنی «کلمه کلیدی، خارج از متن» است.
- نوعی نمایه‌ی گردان است هر بار یکی از کلمات عنوان (به جز کلمات بازدارنده) به ترتیب از عنوان خود خارج شده و مقدم بر سایر اجزای آن قرار می‌گیرد. سپس عنوان به ترتیب طبیعی خود در زیر این واژه یا به دنبال آن می‌آید.
- به آن در اصطلاح نمایه‌ی برون‌بافتی نیز می‌گویند چون مدخل از درون بافت خود خارج شده و مقدم بر سایر اجزا قرار می‌گیرد.
- این نمایه برای جبران مسائل ناشی از محدود بودن جا، کوتاه شدن عنوان و نیز مشکلات خواندنِ نمایه‌ی درون بافتی طراحی شده است.

ارجاع	بافت	مدخل
۱۰۱	اقتصاد مسکن در ایران	اقتصاد
۱۰۳	نقش فرش در اقتصاد ایران	اقتصاد
۱۰۲	بررسی آثار اقتصادی رکود بازار فرش	اقتصادی
۱۰۴	عوامل اقتصادی مهاجرت از روستاها به شهرها	اقتصادی
۱۰۱	اقتصاد مسکن در ایران	ایران
۱۰۳	نقش فرش در اقتصاد ایران	ایران
۱۰۲	بررسی آثار اقتصادی رکود بازار فرش	بازار
۱۰۲	بررسی آثار اقتصادی رکود بازار فرش	رکود
۱۰۴	عوامل اقتصادی مهاجرت از روستاها به شهرها	روستاها
۱۰۴	عوامل اقتصادی مهاجرت از روستاها به شهرها	شهرها
۱۰۲	بررسی آثار اقتصادی رکود بازار فرش	فرش
۱۰۳	بررسی آثار اقتصادی رکود بازار فرش	فرش
۱۰۱	اقتصاد مسکن در ایران	مسکن
۱۰۴	عوامل اقتصادی مهاجرت از روستاها به شهرها	مهاجرت
۱۰۳	نقش فرش در اقتصاد ایران	نقش

۳- نمایه جایگشتی (Permuterm)

- اساس کار آن مانند کوئیک و کووک گردش کلید واژه‌های عنوان است.
- کلید واژه‌های عنوان به صورت دو تایی انتخاب می‌شوند و یک اصطلاح را تشکیل می‌دهند. گردش کار با دو کلید واژه است.
- اصطلاح دوم در محدود کردن اصطلاح اول و اخص کردن آن نقش مهمی دارد.

مثال: با روش Permuterm نمایه‌نامه مقالات بالا را تهیه کنید.

اقتصاد	
ایران	۱۰۱،۱۰۲
فرش	۱۰۳
مسکن	۱۰۱
نقش	۱۰۳
اقتصادی	
آثار	۱۰۲
بازار	۱۰۲
بررسی	۱۰۲
رکود	۱۰۲
روستاها	۱۰۴
شهرها	۱۰۴
عوامل	۱۰۴
فرش	۱۰۲
مهاجرت	۱۰۴
ایران	
اقتصاد	۱۰۳،۱۰۱
فرش	۱۰۳
مسکن	۱۰۱
نقش	۱۰۳
بازار	
آثار	۱۰۲
اقتصادی	۱۰۲
بررسی	۱۰۲
رکود	۱۰۲
فرش	۱۰۲
رکود	
آثار	۱۰۲
اقتصادی	۱۰۲

۱۰۲	بازار	
۱۰۲	بررسی	
۱۰۲	فرش	روستاها
۱۰۴	اقتصادی	
۱۰۴	شهرها	
۱۰۴	عوامل	
۱۰۴	مهاجرت	شهرها
۱۰۴	اقتصادی	
۱۰۴	روستاها	
۱۰۴	عوامل	
۱۰۴	مهاجرت	فرش
۱۰۲	آثار	
۱۰۲	اقتصادی	
۱۰۳	اقتصاد	
۱۰۳	ایران	
۱۰۲	بازار	
۱۰۲	بررسی	
۱۰۲	رکود	
۱۰۳	نقش	مسکن
۱۰۱	اقتصاد	
۱۰۱	ایران	مهاجرت
۱۰۴	اقتصادی	
۱۰۴	روستاها	
۱۰۴	شهرها	
۱۰۴	عوامل	

در این نمایه کلمات: از، به، در، واژه های غیر مجاز و اصطلاحات آثار، بررسی، عوامل، نقش، واژه های نیمه مجاز به حساب می آیند. نکته: واژه های نیمه مجاز به عنوان مدخل اصلی قرار نمی گیرند اما به عنوان بیانگر (یعنی واژه ای دوم) می توانند استفاده شوند.

نکته: اگر در هر عنوانی ۵ کلید واژه وجود داشته باشد، برنامه ی کامپیوتری، ۱۰ جفت اصطلاح می سازد که هر جفت با جابه جا شدن، دو بار در نمایه نامه وارد می شود و در مجموع ۲۰ مدخل را تشکیل می دهد به عبارت دیگر برای هر عنوان $n(n-1)$ مدخل درست می شود که n تعداد کلید واژه های عنوان است.

مثال: عنوان «آموزش ترمیم عکس قدیمی در فتوشاپ» با روش Permuterm چند مدخل خواهد داشت؟

پاسخ: اگر همه کلمات به جز «در» را کلمه کلیدی در نظر بگیریم، تعداد کلمات کلیدی ۵ خواهد بود. بنابراین،

تعداد مدخل‌ها برابر است با: $n(n-1)$ یعنی: $5 \times (5-1) = 20$

- در این نمایه اصطلاح دوم یا اصطلاح همراه، در محدود کردن مفهوم واژه مدخل و اخص شدن معنای آن نقش مؤثری دارد و در واقع همانطور که یک صفت، اسم را توصیف می‌کند این واژه‌ها نیز مدخل را توصیف می‌کنند.

- واژه‌های همراه پس از استخراج از عناوین مربوط، در زیر واژه مدخل به ترتیب الفبایی تنظیم می‌شود.

مثال:

اقتصاد

ایران

مسکن

- در این نمایه‌ها غالباً غیر از واژه‌های بازدارنده یا غیرمجاز (stop list) از یک لیست با عنوان واژه‌های نیمه‌مجاز (semi-stop list) نیز استفاده می‌شود.

نقاط قوت و ضعف نمایه‌ی گردان

- ۱- در مقایسه با نمایه موضوعی بسیار ارزان و ساده و سریع تهیه می‌شود.
- ۲- نمایه گردان منعکس‌کننده واژگان رایج در زمینه مربوطه هستند.
- ۳- نمایه گردان با فراهم آوردن بافت نسبی تصویر روشن‌تری از موضوع می‌دهد.
- ۴- در تهیه این نمایه‌ها، محتوای موضوعی مدرک توسط نمایه‌سازان مختلف تفسیر نمی‌شوند و به این جهت جایی برای اشتباه و اختلاف از لحاظ تعیین و تبیین موضوع باقی نمی‌ماند.
- ۵- گاهی طول مدخل‌های که با یک کلید واژه درست می‌شود به یک یا چند صفحه بالغ می‌گردد.

مثال: برای مقاله‌ای با عنوان آموزش ترمیم عکس قدیمی در فتوشاپ به روش موضوعی و به روش نمایه گردان ۳ کلمه کلیدی بیان کنید؟

به روش نمایه گردان، «در» حذف می‌شود و بقیه کلمات باقی مانده به عنوان کلمات کلیدی هستند. به روش موضوعی، «آموزش فتوشاپ»، «ترفندهای فتوشاپ»، «آموزش فتوشاپ پیشرفته» می‌توانند به عنوان کلیدواژه در نظر گرفته شوند.

انواع نمایه انتهای کتاب

۱- نمایه ساده

در این نمایه:

- مدخل‌ها به ترتیب الفبایی هستند.
- در جلوی هر مدخل جای‌نماها (Locator) وجود دارند.
- فاقد بیانگر است.
- هیچ‌گونه تورفتگی وجود ندارد.

مثال:

کامپیوتر ۴۲، ۴۳

گرافیک ۱۰، ۲۰

۲- نمایه درون بافتی

این نوع نمایه:

- به صورت خطی است.
- به صورت نحوی است.
- به ترتیب پیدایش کلمه در متن است.
- بنابراین الفبایی نیست.

کاربرد: در متون علوم انسانی (ادبیات، الهیات، فلسفه)

مثال) کودکان

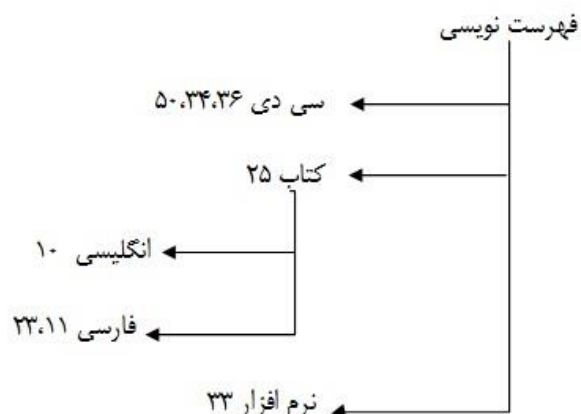
~ استثنای ۱۵، ۲۰؛ ~ تیزهوش ۱۸، ۲۰؛ ~ عقب‌مانده ذهنی ۱۹؛

۳- نمایه برون بافتی

در این نوع نمایه‌ها:

- بیانگرها در زیر مدخل اصلی می‌آیند.
- به صورت الفبایی می‌آیند.
- توالی صفحات در آن‌ها مهم نیست.
- کلماتی مثل حروف اضافه و ربط و ... در آن‌ها استفاده نمی‌شود.
- برخی از بیانگرها در جای خود به عنوان یک مدخل برای بیانگرهای دیگر به حساب می‌آیند.

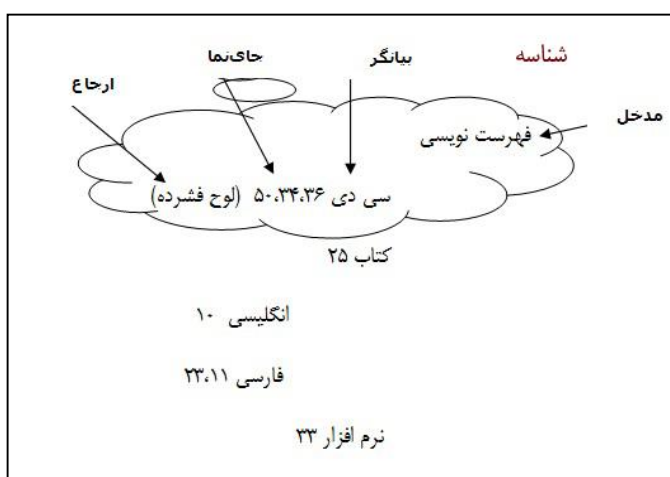
مثال:



تفاوت نمایه انتهایی کتاب، درون بافتی و برون بافتی

- نمایه درون بافتی صورت نحوی دارد.
- یعنی به طور مثال می‌گوییم «رفتار با کودکان» اما در نمایه برون بافتی «با» را نمی‌نویسیم. بنابراین ممکن است بین «رفتار کودکان» و «رفتار با کودکان» ابهام پیش آید.
- درون بافتی برای متون علوم انسانی مناسب است. اما برون بافتی برای متون علمی و فنی و تخصصی.
- نمایه درون بافتی کل‌نگر است.

اجزای نمایه‌های درون بافتی و برون بافتی



شناسه (Heading): به مجموعه‌ی همه اجزای لازم جهت آدرس‌دهی یک سند (یعنی مجموع بخش‌های زیر)، شناسه گفته می‌شود.

مدخل (Entry): کلیت ورود یک کلمه یا نماد به شناسه را گویند. عنصر اصلی یک ترکیب است.

بیانگر (Modifier): نقش محدودکننده‌ی دامنه‌ی معنای مدخل را دارد.
جای‌نما (Locator): محل منابع را نشان می‌دهد.
ارجاع (Reference): برای ارجاع اصطلاح غیر مرجح به مرجح از ارجاع استفاده می‌شود.

اصطلاح مرجح چیست؟

اصطلاحی است که نسبت به مترادف‌هایش برتری (رجحان) دارد.
مثال: دانش به جای علم / رایانه به جای کامپیوتر / یارانه به جای سوبسید

ملاک‌های انتخاب اصطلاح مرجح:

- ۱- رواج داشتن: مثل جامعه‌شناسی به جای علم‌الاجتماع
- ۲- بومی بودن: مثل آرمان‌گرایی به جای ایده‌آلیسم / رایانه به جای کامپیوتر
- ۳- جدید و امروزی بودن: مثل هواپیما به جای طیاره
- ۴- اختصار مشهور به جای عبارت کامل: یونسکو
- ۵- اجتناب از کلمات ترکیبی به جز در مواردی که اصطلاح قابل جداسازی نیست: مثل آبله‌مرغان

توجه: در هر نوع نمایه نسبت به شرایط ممکن است اولویت یکی از موارد بالا بیشتر از دیگری باشد. به طور مثال در یک نوع نمایه تخصصی ممکن است «کامپیوتر» به دلیل رواج داشتن، بر «رایانه» که بومی است اولویت پیدا کند و در یک نوع نمایه، «رایانه» با اولویت بالاتری در نظر گرفته شود که این وظیفه نمایه‌ساز است که نسبت به شرایط، یکی را انتخاب نماید.

گام‌های ایجاد یک نمایه به روش دستی

- گام اول: تعیین و انتخاب خط مشی نمایه‌سازی (برون بافتی یا درون بافتی؟...)
- گام دوم: خواندن سریع متن بدون یادداشت و علامت یا «اریب خوانی» (برای درک کلیت موضوع)
Skim: به خواندن سطحی و بدون دقت روی تک تک کلمات، Skim کردن گفته می‌شود.
Scan: خواندن کلمه به کلمه و ترجمه آن‌ها (در صورت لزوم)
- گام سوم: مشخص کردن کلیدواژه‌ها با خط کشیدن زیر واژه‌ها یا نوشتن در حاشیه‌ی هر صفحه
- گام چهارم: انتقال هر کلید واژه (مدخل) بر روی یک برگه 7.5 در 12.5 سانتی‌متری و ذکر شماره صفحه بر روی هر برگه
- گام پنجم: الفبایی کردن مدخل‌ها و ادغام مدخل‌های تکراری

- گام ششم: ایجاد پیوند مفهومی بین مدخل‌ها؛ به طور مثال ایجاد ارجاع (برای ایجاد روابط می‌توانید از اصطلاح نامه یا فرهنگ لغت کمک بگیرید)

نکاتی در مورد ساختار زبان شناسی مدخل

- ۱- فقط اسم یا عبارت اسمی می‌تواند به عنوان مدخل قرار گیرد. یعنی جزء اصلی مدخل حتماً باید اسم باشد ولی بیانگر می‌تواند صفت یا مضاف الیه باشد.
مثال: انگشتر طلا (صحيح است چون «انگشتر» که یک اسم است به عنوان مدخل و «طلا» که یک صفت است به عنوان بیانگر قرار گرفته)
- سپید دندان (غلط است چون «سپید» که صفت است به عنوان مدخل و «دندان» که اسم است به عنوان بیانگر قرار گرفته)
- ۲- صفت و قید نمی‌توانند مدخل قرار گیرند مگر اینکه جانشین اسم شده باشند.
مثال: بالا بر (صحيح است چون صفت «بالا» در اینجا در جایگاه اسم قرار گرفته)
- ۳- وارونه‌سازی ممنوع (به جز اسامی افراد)
مثال: استثنایی، کودکان (غلط است)
خوارزمی، محمد (صحيح است)

روال تهیه یا خط مشی نمایه

- تعداد متوسط کلید واژه‌ها مشخص شود.
- نوع نمایه باید تعیین شود. (درون بافتی یا برون بافتی)
- زبان نمایه‌سازی مشخص شود. (آزاد، کنترل شده، ترکیبی)
- بیانگرها باید مشخص باشند. (جمع یا مفرد، تکلیف کلمات خارجی مشخص شود)
- نوع بازیابی تعیین شود.
- سیاست تعیین کلمات مرجح یا نامرّجح مشخص شود.
- ارجاعات و جای‌نماها مشخص شوند.
- انواع نمایه مشخص شوند (موضوع؟ نویسندگان؟...)
- استانداردهای با استفاده از ابزارهای موجود
- خودکارسازی نمایه با استفاده از نرم‌افزارها
- به‌روزرسانی نمایه‌ها

معیارهای ارزیابی نمایه

نمایه‌سازان نمایه‌ها را در سه سطح مورد ارزیابی قرار می‌دهند:

۱- سطح فنی یا **Technical**:

در این سطح نمایه باید دارای زبان مناسب و شکل قابل درک باشد و به سادگی از آن استفاده کرد.

۲- سطح معنایی یا **Semantic**:

در این سطح واژه‌ها باید معانی را بدون ابهام منتقل کند.

۳- سطح میزان کارایی یا **Effectiveness**:

در این سطح نمایه باید اطلاعات مرتبط را به درستی شناسایی کند و در بازیابی اطلاعات مؤثر باشد.

از نظر آقای کلوند (Cleveland) در ارزیابی نمایه‌ها چه نکاتی باید مد نظر قرار گیرد؟

- ۱- دامنه تحت پوشش نمایه کامل باشد.
- ۲- در انتخاب اصطلاح انسجام داشته باشد.
- ۳- اصطلاحات انتخاب شده با سطح کاربران متناسب باشد.
- ۴- ارجاعات به میزان کافی وجود داشته باشد.
- ۵- زنجیره‌های بسیار طولانی از سرعنوان‌های فرعی به هم پیوسته، وجود نداشته باشد.
- ۶- سرعنوان‌های فرعی به درستی سرعنوان‌های اصلی را نشان دهد.
- ۷- مکان‌نمای نادرست وجود نداشته باشد.
- ۸- زنجیره بسیار طولانی از مکان‌نماها وجود نداشته باشد. یک نمایه خوب نباید بیش از ۱۰ الی ۱۲ ارجاع برای هر اصطلاح داشته باشد.
- ۹- نظم الفبایی در متن، یکپارچه و صحیح باشد.
- ۱۰- اشتباه املایی نداشته باشد.
- ۱۱- ارجاع نادرست و دو طرفه وجود نداشته باشد.

ویژگی‌های یک نمایه خوب چیست؟

- ۱- **صحت:** هرچه خطاهای ارجاع به صفحه‌ها کاهش یابد و از مدخل‌های صحیح استفاده شود، نمایه کارایی بیشتری خواهد داشت.
- ۲- **عمق نمایه‌سازی:** اینکه در هر صفحه از سند چند کلمه به عنوان مدخل قرار گرفته، عمق نمایه‌سازی گفته می‌شود. عمق نمایه‌سازی باید با نیاز کاربران متناسب باشد. به عنوان نمونه، یک نمایه‌ی متوسط ۳ تا ۵ مدخل اصلی برای هر صفحه از متن و یک نمایه‌ی ضعیف ۱ تا ۲ مدخل دارد.
- ۳- **ایجاز:** هرچه از سرعنوان‌های فرعی به صورت فشرده و به هم بسته استفاده شود نمایه بهتری خواهیم داشت.
- ۴- **ارجاعات متقابل:** ارجاعات باید صحیح، منطقی و مفید باشد.
- ۵- **سرعنوان‌های منطقی:** سرعنوان‌ها باید از میان گزینه‌های مختلف بهترین آن‌ها انتخاب شده باشد.

عوامل مؤثر در نمایه‌سازی

فرایند باز یابی	فرایند نمایه‌سازی	واژگان	مدرک	کاربر	نمایه‌ساز
- هزینه	- جامعیت	- کیفیت	- محتوای موضوعی	- نیازهای	- دانش موضوعی
- زمان	- مانعیت	- مدخل‌ها	- حجم	- اطلاعاتی	- تجربه
- میزان ریزش	- صحت	- کیفیت ساختار	- زبان	- میزان تلاش	- تلاش فکری
- کارایی نظام	- کل‌نگری	- عدم ابهام		- شیوه	
- شکل فیزیکی	- جزئ‌نگری	- توانایی		- جست‌وجو	
- برون‌داد	- فنون نمایه‌سازی	- روزآمدسازی			

گنج‌واژه یا اصطلاح‌نامه

مجموعه اصطلاحات یک رشته است که میان آن‌ها روابط معنایی، رده‌ای، و سلسله مراتبی برقرار شده و توانایی آن را دارد که موضوع آن رشته را با همه جوانب اصلی و فرعی و وابسته، به‌گونه‌ای نظام‌یافته و به منظور ذخیره و بازیابی اطلاعات ارائه دهد. [معادل انگلیسی گنج‌واژه، کلمه Thesaurus است که ریشه در زبان لاتینی باستان دارد و به معنای گنجینه، ذخیره و مجموعه به کار می‌رود].

برخی اصطلاح‌نامه‌های مشهور: Encarta, Encyclopedia (دانشنامه مایکروسافت)، Microsoft Computer Dictionary

اهداف اصطلاح‌نامه

- ۱- نمایاندن یک ساختار مشخص برای زمینه‌ی خاصی از دانش به طوری که هم نمایه‌ساز و هم جست‌وجوگر بتوانند از گستره آن زمینه و ارتباط میان مفاهیم آن آگاهی یابند.
- ۲- ارائه اصطلاحات استاندارد در زمینه معین
- ۳- برقراری نظام ارجاعات میان اصطلاحات و رده‌بندی اصطلاحات به صورت سلسله مراتبی
- ۴- تأکید بر توجه به نیازهای اطلاعاتی استفاده‌کننده
- ۵- تعیین اصطلاحات مجاز و مشخص کردن حدود معانی اصطلاحات به منظور ایجاد هماهنگی در نمایه‌سازی

روابط میان اصطلاحات

- ۱- رابطه هم ارزی یا Equivalence Relation
میان اصطلاح پذیرفته‌شده و اصطلاح پذیرفته‌نشده این ارتباط برقرار می‌شود.
مثال: اصطلاح «رایانه» به جای «کامپیوتر»، یا «لوح فشرده» به جای سی.دی
- ۲- رابطه سلسله مراتبی یا Heirarchcal Relation
بیان‌کننده رابطه اعم و اخص میان مفاهیم است که در واقع، اصطلاح‌نامه‌ها را از واژه‌نامه‌های متداول متمایز می‌کند.
مثال: رابطه صفحه نمایش و LCD, LED, CRT
- ۳- رابطه همبسته یا Associative Relation
رابطه میان دو اصطلاح که به دلیل وابستگی معنایی، وجود یکی دیگری را نیز به ذهن متبادر می‌کند.
مثال: اصطلاح اسب و سوارکاری

انواع اصطلاح

۱- اصطلاح اعم (ا.ع.) یا Brother Term

اصطلاح عام تر نسبت به اصطلاح فعلی

دو مثال:

انگور ا.ع. : میوه	توپولوژی ا.ع. : شبکه
----------------------	-------------------------

۲- اصطلاح اخص (ا.خ.) یا Narrow Term

اصطلاح خاص تر نسبت به اصطلاح فعلی

مثال:

میوه ا.خ. : انگور	ادوبی ا.خ. : فتوشاپ
----------------------	------------------------

۳- اصطلاح وابسته (ا.و.) یا Related Term

اصطلاحی که به اصطلاح فعلی به نحوی ارتباط دارد.

مثال:

انگور ا.و. : شیره	شبکه ا.و. : مودم
----------------------	---------------------

۴- اصطلاح به کار ببرید (ب.ک.) یا Use

ارجاع از اصطلاح نامرئج به مرئج

مثال:

تعلیم و تربیت ب.ک. : آموزش و پرورش	ایمیل ب.ک. : رایانامه
---------------------------------------	--------------------------

۵- اصطلاح به کار ببرید به جای (ب.ج.) یا Used for

ارجاع از اصطلاح مرئج به اصطلاح نامرئج

مثال:

آموزش و پرورش ب.ج. : تعلیم و تربیت	موشواره ب.ج. : موس
---------------------------------------	-----------------------

۶- اصطلاح راس (ا.ر.) یا Top Term

اصطلاح سرمقوله‌ی مجموعه‌ای از اصطلاحات می‌باشد.

مثال:

میوه	LED
ا.ر.: خوراکی‌ها	ا.ر.: Monitor

۷- یادداشت دامنه (ی.د.) یا Scope Note

یادداشتی برای توضیح مدخلی که دارای ابهامی و یا دارای چند معنا است و یا معنی خیلی کلی دارد استفاده می‌شود.

مثال:

فرهنگ	شبکه
ی.د.: آداب و رسوم	ی.د.: شبکه‌های کامپیوتری

انواع رابطه سلسه مراتبی

۱- رابطه جنس و نوع یا Generic

این رابطه دقیق‌ترین و قطعی‌ترین رابطه سلسله مراتبی است که با کلمات «همه» برای «جنس» و «بعضی» برای «نوع» قابل شناسایی است.

مثال: رابطه میان کلمه «نشخوارکنندگان» (جنس) و کلمات «بز» (نوع)، «گاو» (نوع) و «گوسفند» (نوع) همه بزها نشخوارکننده هستند. بعضی از نشخوارکنندگان بز هستند.

مثال: رابطه بین «پردازشگر» و Intel و AMD

۲- رابطه کل و جزء یا Whole-part

نوعی رابطه اعم و اخص است.

مثال: رابطه بین کلمه «دست» (جزء) و «بدن» (کل) / رابطه بین CPU (کل) و ALU (جزء)، CU (جزء)

۳- رابطه مصداقی یا Instance Relation

این ارتباط بیشتر در مورد اسامی خاص (حافظ، دماوند و...) کاربرد دارد و ارتباط میان جنس و نوع به حساب نمی‌آید بلکه بیانگر یک مثال و مصداق است.

مثال: دماوند (مصداق یک قله) / LG (مصداق مانیتور)

تفاوت واژه نامه و اصطلاح نامه

واژه نامه‌ها نظم الفبایی دارد و به دنبال هم می‌آیند، ولی با هم مرتبط نبود و نسبت به هم حالت خنثی دارند. اما اصطلاح‌نامه معمولاً نظم الفبایی-رده‌ای دارد و کلمات با هم مرتبط بوده و روابط معنایی دارند.

چکیده یا Abstract

An abstract is a brief summary of a research article thesis, conference, proceeding or any in-depth analysis of a particular subject or discipline, and is often used to help the reader quickly ascertain the papers purpose.

یک چکیده یک خلاصه کوتاه از یک مقاله تحقیقاتی از پایان‌نامه، نقد، مقاله ارائه شده در یک کنفرانس، یا هر تحقیق عمیق از یک موضوع یا منظومه است و اغلب برای کمک به خواننده برای فهم سریع هدف مطلب استفاده می‌شود.

تعریف دیگر: خلاصه‌ای است از یک نوشته که شامل فشرده‌ی تمام مطالب مهم یا فشرده قسمت‌های ویژه آن نوشته می‌باشد.

گاهی شامل اصطلاحات و کلید واژه‌های متن است.

تعریف چکیده‌نویسی Abstracting

به عملیات تهیه چکیده از یک نوشته در اصطلاح «چکیده‌نویسی» می‌گویند.

چکیده به ارزش سند می‌افزاید. چرا؟

یکی از دلایل آن این است که از روی چکیده می‌توان فهمید آنچه نویسنده در سند گفته است برای خود نویسنده قابل فهم بوده یا نه.

توجه: انتهای هر فصل پروژه یا پایان‌نامه یا رساله یا مقاله یا هر نوع مستندات باید در حد چند پاراگراف چکیده‌ای از آن فصل بیان شود.

دو فرایند اصلی در تهیه چکیده

۱- هرس متن یا حذف اضافات:

در این فرایند قسمت غیرضروری حذف شده و نکات مهم و ضروری حفظ می‌شوند.

۲- عصاره‌گیری متن:

در این فرایند از روی متون باقیمانده، به اصل مفهوم محتوای مدرک می‌رسیم.

انواع چکیده بر حسب لحن

۱- چکیده تمام‌نما یا Informative Abstract

این نوع چکیده تا حد ممکن اطلاعات روشنی از مباحث کمی و کیفی مندرج در مدرک ارائه می‌کند بنابراین کمی طولانی‌تر از سایر انواع چکیده است.

طول این نوع چکیده باید متناسب با نوع و محتوای مدرک باشد اما غالباً در مقالات چاپ شده در مجلات بین ۱۰۰ تا ۲۵۰ لغت و در پایان‌نامه‌ها نهایتاً ۵۰۰ لغت در نظر گرفته می‌شود. تهیه این نوع چکیده معمولاً زمان‌بر و پرهزینه است و نیاز به مهارت خاص و کافی دارد.

۲- چکیده راهنما یا **Indicative Abstract**

چکیده‌ای است که اطلاعات کامل را نمی‌دهد بلکه به یافته‌ها هدایت می‌کند و بیشتر تأکید روی نتایج مدرک است.

هدف اصلی آن ایجاد انگیزه برای مراجعه به متن است.

۳- چکیده تلفیقی یا **Compound Abstract**

ترکیبی از چکیده راهنما و تمام‌نما است.

با توجه به اینکه در چکیده تمام‌نما امکان بیان و نمایش جداول و نمودارهای مدرک نیست می‌توان از چکیده راهنما برای رفع این کمبود استفاده کرد.

۴- چکیده انتقادی یا **Critical Abstract**

علاوه بر معرفی محتوای مدرک، به ارزیابی آن هم می‌پردازد.

به ندرت از این نوع چکیده استفاده می‌شود چون تهیه آن نیاز به مهارت در چکیده‌نویسی و داشتن تخصص موضوعی در سطح بالا دارد.

در این نوع چکیده، چکیده‌نویس عقاید و نظرات خود را بیان می‌کند.

انواع چکیده بر حسب حجم

۱- چکیده تلگرافی یا **Telegraphic Abstract**

خیلی کوتاه و در حد یک یا یک و نیم جمله است.

این نوع چکیده در وب سایت‌ها برای بیان خلاصه کوتاهی از آموزش‌ها و مقالات سایت استفاده می‌شود.

مثال:

چگونه فایل فارسی که به صورت ناخوانا (خرچنگ قورباغه!!) ذخیره شده را اصلاح کنیم؟
گاهی یک فایل فارسی به دست شما می‌رسد که وقتی باز می‌کنید، کاراکترهایی شبیه به این می‌بینید: Ó ÇÒ Åä ÍÇÄ ÇNOCÁ äi Oæel. چگونه این‌ها را اصلاح می‌کنید؟
نویسنده: Hamid خوانده شده: 2949

روانشناسی رنگ‌ها

نویسنده: خوانده شده: 2966

آشنایی با انواع قفل‌های نرم‌افزاری برای CDها
یکی از مهمترین مسائل، پس از تولید یک محصول، قفل‌گذاری بر روی آن است! مقاله‌ای خوب برای آشنایی با انواع قفل‌های نرم‌افزاری...
نویسنده: Hamid خوانده شده: 2329

نرم افزار windjview

این نرم افزار برای باز کردن فایلها (کتابها و مقالات و ...) با پسوند windj بکار می‌رود...
نویسنده: sahel20 خوانده شده: 4148

آموزش گام به گام استفاده از Fdisk برای پارتیشن‌بندی هارد دیسک (مقاله‌ای کامل و جامع)

فرض کنید همین امروز یک هارد دیسک جدید خریده‌اید، طبیعتاً اولین کاری که باید انجام دهید، در اصطلاح، پارتیشن بندی هارد دیسک است. برنامه Fdisk در این زمینه به شما کمک می‌کند...
نویسنده: Hamid خوانده شده: 4092

۲- چکیده عنوانی یا Titular Abstract

با توجه به اینکه همیشه عنوان، بیانگر متن نیست، یک عنوان ساختگی ایجاد می‌شود که جایگزین عنوان مبهم باشد.

به طور مثال عنوان یک مقاله عبارت است از: در هیروشیما چه گذشت؟ (معرفی مستند Hiroshima)

۳- چکیده عام یا General Abstract

نسبتاً طولانی است و تاکنون استاندارد دی برای آن تهیه نشده است. بنابراین بستگی به سیاست پایگاه یا مؤسسه یا نشریه دارد.

انواع چکیده بر حسب تهیه کننده

۱- چکیده مؤلف یا Author Abstract

مؤسسات ترجیح می‌دهند که خود مؤلف که بر متن مسلط است آن را بنویسد.

۲- چکیده حرفه‌ای یا تخصصی یا Professional Abstract

توسط افراد حرفه‌ای نوشته می‌شوند و نسبت به چکیده مؤلف دارای نفوذ و اهمیت بیشتری است.

۳- چکیده نویسی دستی

این نوع چکیده توسط انسان و به صورت دستی تهیه می‌شود.

۴- چکیده نویسی ماشینی

تمام کار به وسیله کامپیوتر است. برخی از الگوریتم‌های آن شامل نظریه گراف‌ها، روش ارزیابی Rouge و

Key method

نمونه سایت‌های چکیده‌نویسی ماشینی

- سایت ایجاز <http://ejaz.um.ac.ir>
- سایت تحقیقات علوم اسلامی نور <http://textmining.noorsoft.org/Fa/summarization>
- سایت <http://tools4noobs.com/summarize> (با قابلیت پشتیبانی از زبان فارسی)
- سایت <http://autosummarizer.com> (بدون قابلیت پشتیبانی از زبان فارسی)

نکات مهم در تهیه چکیده مقالات و پایان‌نامه‌ها و طرح‌های پژوهشی

- شروع با یک بحث نظری در یک یا دو خط
- مسأله‌ی تحقیق در دو یا سه خط
- اهداف تحقیق و فرضیه‌های پژوهش
- روش تحقیق و ابزارهای پژوهش، جامعه پژوهش
- یافته‌ها و نتایج
- بیشتر از یک و نیم صفحه نباشد (حدود ۳۰۰ تا ۵۰۰ واژه)

گام‌های تهیه چکیده

- ابتدا تعیین خط مشی
- یادآوری این نکته که چکیده، حاوی خلاصه‌ی کل مدرک است.
- تعیین اهداف چکیده‌نویسی
- چکیده باید حاوی کلیدواژه‌های متن اصلی است.
- تعیین نوع و انواع چکیده مورد استفاده در چکیده‌نویسی
- تعیین سطح علمی و مهارتی چکیده‌نویس یا چکیده‌نویسان
- تعیین این که چکیده‌های تهیه شده ویراستاری شده‌اند یا نه.
- تعیین نوع حروف‌چینی، صفحه‌آرایی و مسائلی از این قبیل
- تعیین محل درج اطلاعات کتابشناختی در ابتدا یا انتهای چکیده
- تعیین جملاتی که نباید در چکیده به کار روند.
- رعایت نظم و ترتیب متن اصلی چکیده

رده‌بندی چیست؟

رده‌بندی روشی برای دسته‌بندی کتب و دانش است و گروهی که دارای ویژگی‌های یکسان باشد رده نامیده می‌شود.

انواع رده‌بندی

۱- رده‌بندی دیوئی (Dewey-Decimal-Classification)

یک رده‌بندی مشهور در علوم کتابداری و اطلاع‌رسانی برای تعیین موضوع اصلی کتاب می‌باشد. این رده‌بندی جهانی در سهولت و دقت انتخاب کتاب، تعیین مکان آن در کتابخانه‌ها و فهرست‌نویسی علمی کتاب‌ها نقش مؤثر دارد.

در این رده‌بندی، همه دانش‌های بشری ابتدا به ۱۰ رده یا موضوع کلی تقسیم می‌شود و هر رده در نوبت دوم به ۱۰ رده دیگر و در نوبت سوم هر رده به ۱۰ رده دیگر تقسیم می‌شود. در مجموع ۱۰۰۰ عنوان موضوعی قابل تصور است.

۱۰ رده اصلی عبارتند از:

۰۰۰ = کلیات

۱۰۰ = فلسفه و روانشناسی

۲۰۰ = دین

۳۰۰ = اجتماع

۴۰۰ = زبان

۵۰۰ = علوم

۶۰۰ = فنون

۷۰۰ = هنر، سرگرمی و ورزش

۸۰۰ = ادبیات

۹۰۰ = تاریخ و جغرافیا

۲- رده‌بندی کنگره

روش برای رده‌بندی موضوع کتب و دانش‌هاست و به‌جای شماره‌گذاری، یک یا چند حرف انگلیسی برای تعیین موضوعات کلی در نظر گرفته و از اعداد برای توسعه آن‌ها بهره می‌گیرد.

این رده‌بندی اساس نظام رده‌بندی کتابخانه کنگره آمریکا می‌باشد. این طرح متشکل از ۲۱ رده در قالب حروف الفبای انگلیسی می‌باشد و در حال حاضر از ۲۶ حرف الفبا حروف V, X, W, O, I در این رده بندی به کار نمی‌رود و برای گسترش‌های آینده ذخیره شده‌اند. این نظام، موضوعی و ساختار آن به صورت سلسله مراتبی از کل به جز

است و حروف بزرگ در کنار رده اصلی برای رده‌های فرعی به کار می‌رود. همچنین برای تقسیمات فرعی از اعداد صحیح استفاده می‌شود.

به طور مثال رده‌های اصلی ۲۱ گانه دانش عبارتند از:

- ۱- A کلیات
- ۲- B فلسفه، روانشناسی، مذهب
- ۳- C علوم وابسته به تاریخ
- ۴- D تاریخ
- ۵- E تاریخ ایالات متحده
- ۶- F تاریخ محلی ایالات متحده؛ تاریخ آمریکا
- ۷- G جغرافیا، مردم شناسی، سرگرمی‌ها
- ۸- H علوم اجتماعی
- ۹- J علوم سیاسی
- ۱۰- K حقوق
- ۱۱- L آموزش و پرورش
- ۱۲- M موسیقی
- ۱۳- N هنرهای تجسمی
- ۱۴- P زبان و ادبیات
- ۱۵- Q علوم طبیعی و ریاضی
- ۱۶- R پزشکی
- ۱۷- S کشاورزی
- ۱۸- T تکنولوژی
- ۱۹- U علوم نظامی
- ۲۰- V علوم دریایی
- ۲۱- Z کتابشناسی و کتابداری

۳- رده‌بندی با کمک کد شابک یا (International Standard Book Number) ISBN:

شابک، شماره استاندارد بین‌المللی کتاب است. این شماره در ابتدا کدی ۹ رقمی بود که در سال ۱۹۶۷ توسط آقای گوران فاستر ابداع شد و پس از آن در سال ۱۹۷۰ به ده رقم تغییر یافت و نهایتاً از سال ۲۰۰۷ تاکنون یک کد ۱۳ رقمی می‌باشد.

شابک ۱۳ رقمی دارای ۵ بخش است که هر بخش با یک - از هم جدا می‌شود:

- بخش اول از سمت چپ ۹۷۸ یا ۹۷۹ است.
- بخش دوم کد کشور مربوط به کتاب است. (به طور مثال کد ایران ۹۶۴ است)

- بخش سوم کد خاص انتشارات کتاب است.
- رقم بعدی کد خاص آن کتاب در انتشارات است.
- بخش آخر در اصطلاح Checksum ارقام دیگر است.

Checksum: روشی برای تشخیص خطا در انتقال داده از مبداء به مقصد است.

روش به دست آوردن Checksum در شابک ۱۳ رقمی:

مثال: در کد شابک زیر، رقم checksum را به دست آورید:

۹۷۸-۹۶۴-۵۳۱-۰۸۳-....

پاسخ:

- از سمت چپ، ارقام را یک در میان در ۱ و ۳ ضرب کنید.
- نتیجه را با هم جمع کرده و بر ۱۰ تقسیم کنید.
- باقیمانده را از ۱۰ کم کنید تا Checksum به دست آید.

$$9*1 + 7*3 + 8*1 + 9*3 + 6*1 + 4*3 + 5*1 + 3*3 + 1*1 + 0*3 + 8*1 + 3*3 = 115$$

$$115 \text{ mod } 10 = 5$$

$$10 - 5 = 5$$

روش‌های کمک گرفتن از یک کتابدار (مفهوم Ask a Librarian)

۱- سایت کتابدار:

در این روش با مراجعه به سایت کتابدار از طریق پر کردن فرمی از یک کتابدار برای یافتن کتب و مقالات مرتبط با تحقیق خود سؤال می‌پرسید و کتابدار از طریق ایمیل شما را به کتاب‌های کتابخانه خود یا دیگر کتابخانه‌ها ارجاع می‌دهد.

نمونه‌هایی از سایت‌های ارائه‌دهنده خدمات فوق:

- سایت کتابخانه کنگره آمریکا (loc.gov): این کتابخانه، بزرگ‌ترین کتابخانه دنیاست و در واشنگتن‌دی‌سی قرار گرفته است.

- SPL.ORG

- نمونه‌ی ایرانی: کتابخانه ملی ایران (nlai.ir)

The screenshot shows the homepage of The Seattle Public Library. At the top, there is a navigation bar with links for Home, Audiences, Using the Library, Library Collection, Calendar of Events, Locations, and About the Library. The date August 19, 2013 is displayed. Below the navigation bar, there is a search bar and a 'LIBRARY LOCATOR' section. The main content area is divided into several sections: 'LIBRARY NEWS AND EVENTS' with an RSS feed, 'LIBRARIES FOR ALL' with a list of languages, and 'AUDIENCES' with a list of age groups. A red arrow points to the 'Need Help? Ask a Librarian 24/7' button in the bottom right corner.

۲- چت زنده:

در این روش که کتابخانه‌های بزرگ از آن پشتیبانی می‌کنند، کتابداران به صورت آنلاین آماده پاسخگویی و راهنمایی کاربران هستند.

۳- SMS و تماس تلفنی:

برخی کتابخانه‌ها یک شماره تماس و یک شماره پیامک ارائه کرده‌اند که می‌توانید سؤال خود را از آن طریق بپرسید. مانند کتابخانه دانشگاه Yale (ask.library.yale.edu) یا دانشگاه سانتاکروز

(<http://guides.library.ucsc.edu/ask-a-librarian/phone>)

۴- پرسش از طریق ایمیل:

برخی کتابخانه‌ها یک آدرس عمومی ارائه کرده‌اند که می‌توانید سؤال خود را از آن طریق بپرسید.

The screenshot shows the top navigation bar of the Library of Congress website. The logo 'LIBRARY OF CONGRESS' is on the left. To its right are three buttons: 'ASK A LIBRARIAN', 'DIGITAL COLLECTIONS', and 'LIBRARY CATALOGS'. A red arrow points to the 'ASK A LIBRARIAN' button. Below the navigation bar, the page is divided into two main sections. The left section is titled 'About the Library' and features a large banner for the 'NATIONAL BOOK FESTIVAL' on 'SEPT 21-22'. Below the banner are five circular icons and a play button. Underneath are links for 'The Library's Mission', 'The History of the Library', 'Awards & Honors', and 'Jobs & Fellowships'. The right section is titled 'Search' and includes a search bar with 'All Formats' and 'Search Loc.gov' options. Below the search bar are 'Print', 'Subscribe', and 'Share' icons. The 'Collection Highlights' section follows, displaying a grid of six featured items: 'American Memory', 'Prints & Photographs', 'Historic Newspapers', and three other items with book covers or historical photos.

آنتولوژی چیست؟

آنتولوژی در لغت از دو کلمه **onto** به معنی «هستی» و **logy** به معنی «مطالعه» تشکیل شده است. ریشه آنتولوژی در فلسفه است و به ارسطو نسبت داده می‌شود. در علوم کامپیوتر و در حوزه وب معنایی، آنتولوژی مفهوم کلمات و ارتباط بین آن‌ها در حوزه‌ای که مورد استفاده قرار می‌گیرند را نشان می‌دهد. امروزه با حجم عظیمی از اطلاعات مواجه هستیم. به منظور دستیابی مؤثر به اطلاعات، به سیستمی نیاز است که بتواند به طور مناسبی، اطلاعات را از وب استخراج کرده و به طور مناسبی به کاربران ارائه دهد، اما این کار به سادگی میسر نمی‌باشد. آنتولوژی به وجود آمده است تا این مشکلات را مرتفع سازد.

انواع وب

۱- وب ۱ (وب سنتی) (Traditional Web)

این وب همان وب اولیه است که توسط آقای تیم برنرزی ابداع شد. حاوی صفحات ساده‌ی بدون جذابیت و بدون هوشمندی.

۲- وب ۲ (وب اشتراکی) (Social Web)

در این دوران سایت‌های اشتراک جمعی یا شبکه‌های اجتماعی (Social Networks) مثل فیس‌بوک، توئیتر، اینستاگرام و... ظهور کردند و کاربران توانستند داده‌های خود را به راحتی با یکدیگر به اشتراک بگذارند و بنابراین دسترسی به داده‌ها آسان‌تر شد.

۳- وب ۳ (وب معنایی یا وب هوشمند) (Semantic Web)

در این وب پیش‌بینی می‌شود صفحات وب بیش از گذشته برای موتورهای جستجو به طور کامل شناخته شده باشند، بنابراین نسبت به نیاز کاربر بهترین نتایج یافته و نمایش داده خواهد شد.

روش‌های معنا بخشیدن به صفحات وب

۱- XML

۲- URI

۳- RDF

XML چیست؟

XML مخفف Extensible Markup Language و به معنی «زبان نشانه‌گذاری قابل توسعه» است.

فایل XML یک فایل متنی ساده و یک پایگاه داده قابل حمل به حساب می‌آید.

در این زبان تگ‌ها توسط برنامه‌نویس تعریف می‌شود و خود زبان، تگ‌های مشخصی ندارد. (پس یعنی دیگران هستند که زبان را «توسعه» می‌دهند. پس این، یک «زبان قابل توسعه» است) معمولاً هر تگ والد، نماد یک موجودیت و تگ‌های داخلی، صفات آن موجودیت یا همان سرستون‌ها در پایگاه داده هستند.

این زبان امروزه در تمامی نرم‌افزارها برای انتقال اطلاعات یا تنظیمات از یک محل به محل دیگر کاربرد دارد.

مثال: جدول زیر را به زبان XML تبدیل کنید؟

books

A1	A2	A3	A4
E1	E2	E3	E4
D1	D2	D3	D4
S1	S2	S3	S4

<BOOKS>

<BOOK>

<A1>E1</A1>

<A2>E2</A2>

<A3>E3</A3>

<A4>E4</A4>

</BOOK>

<BOOK>

<A1>D1</A1>

<A2>D2</A2>

<A3>D3</A3>

<A4>D4</A4>

</BOOK>

<BOOK>

<A1>S1</A1>

<A2>S2</A2>

<A3>S3</A3>

<A4>S4</A4>

</BOOK>

</BOOKS>

مثال: جداول موجود در پایگاه داده‌ی یک کتابخانه به شکل زیر است، از داده‌های آن یک خروجی XML بگیرید.

Books

bid	Code	ISBN	title	author	publisher	pubDate
1	1001	978-964-531-102-3	Learn Internet	Hamid Reza Niroomand	Aftabgardan	2015
2	1002	978-964-531-302-1	C# in 24 Hours	Qomi	Naqous	2010

Members

mid	Code	FirstName	LastName
1	100	Ali	Hasani
2	101	Hojjat	Rahimi

```

<library>
  <books>
    <book>
      <bid>1</bid>
      <code>1001</code>
      <ISBN>978-964-531-102-3</ISBN>
      <title>Learn Internet</title>
      <author>Hamid Reza Niroomand</ author>
      < publisher >Aftabgardan</ publisher >
      <author>2015</ author>
    </book>
    <book>
      <bid>2</bid>
      <code>1002</code>
      <ISBN>978-964-531-302-1</ISBN>
      <title>C# in 24 Hours</title>
      <author>Qomi</ author>
      < publisher >Naqous</ publisher >
      <author>2010</ author>
    </book>
  </books>
  <members>
    <member>
      <mid>..</mid>
      به همین صورت، مانند بالا ادامه دهید...
    </member>
  </members>
  ...
</library>

```

URI چیست؟

مخفف Uniform Resource Identifier و به معنی «شناسه یک‌شکل منابع» و شناسه یا آدرسی است که از روی اجزای آن می‌توان اطلاعات کافی در مورد آن منبع از جمله رده یک شی و رده های بالاتر آن به دست آورد.

مثالی از یک URL بدون به‌کارگیری مفاهیم وب معنایی:

<http://aftab.cc/linkestan/showlink.php?id=1400>

از روی این آدرس نمی‌توان فهمید که این لینک در مورد چه چیزی است.

همان لینک با به‌کارگیری مفاهیم وب معنایی:

<http://aftab.cc/linkestan/link/pdf/learn-c-sharp-in-24-hours>

از روی این لینک می‌توان فهمید که لینک به یک pdf در مورد آموزش C# است.

URL چیست؟

مخفف Uniform Resource Locator و به معنی «مشخص‌کننده‌ی موقعیت منبع به صورت یک‌شکل» است که یک نوع URI به شمار می‌آید. منظور از URL همان آدرس عکس‌ها و ویدئوها و صفحات روی اینترنت است.

RDF چیست؟

مخفف Resource Description Framework و به معنی «چهارچوب توصیف منابع» است.

زبانی است که به لغات یک صفحه برای موتورهای جستجو معنا می‌بخشد و ارتباط بین لغات را مشخص می‌کند. با استفاده از XML می‌توانیم موجودیت‌های مختلف و صفات آن‌ها را تک‌گذاری کنیم اما معنای آن‌ها را نمی‌توانیم بیان کنیم به عنوان مثال: اگر کسی برچسب قیمت را با PRICE نشانه‌گذاری کند و شخص دیگری با COST نشانه‌گذاری کند هیچ راهی برای فهم مترادف بودن این دو عبارت به وسیله XML وجود ندارد، اینجاست که RDF به کار می‌آید.

هر عبارت RDF به صورت یک سه-تایی (Tripple) بیان می‌شود: فاعل، گزاره، مفعول.

۳ موضوع اصلی در RDF

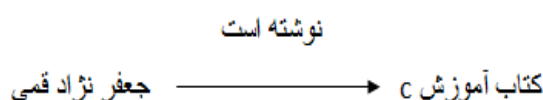


۱- Subjects: فاعل جمله

۲- Predicates: صفت یا فعلی که فاعل آن را انجام می‌دهد.

۳- Objects: مفعول یا اسمی که جانشین مفعول شده است و کار، روی آن انجام می‌شود.

مثال:



مزایای RDF نسبت به XML:

۱- RDF یک فهم مشترک از یک مفهوم واحد را ارائه می‌دهد زیرا گاهی اوقات ما برای یک مفهوم از دو لغت

متمایز استفاده می‌کنیم یا بلعکس از یک لغت برای دو مفهوم متفاوت استفاده می‌کنیم.

۲- RDF بین مفاهیم در وب و دنیای واقعی ارتباط برقرار می‌کند.

مثال: جمله «آقای تیم برنزی وب را اختراع کرد» را به صورت RDF نمایش دهید.

<وب#><اختراع کرد#><آقای تیم برنزی#>

روش اعمال RDF به یک صفحه با HTML5

مثال: عبارت [نام من حمید رضا نیرومند است. من ایران را دوست دارم. کتاب مورد علاقه من «آینده خود را خلق کنید» نوشته «برایان تریسی» است] را به صورت HTML5+RDFa بنویسید؟

```
<html prefix="dc: http://purl.org/dc/elements/1.1/" lang="fa">
  <head>
    <title>سایت شخصی حمید رضا نیرومند</title>
  </head>
  <body about="http://niroomand.ir">
    <h1>صفحه اصلی</h1>
    <p>است و من <span property="foaf:nick">حمید رضا نیرومند</span> نام من<span property="foaf:nick">حمید رضا نیرومند</span>
    <a href="http://www.iran.ir/" rel="foaf:interest" lang="fa">ایران</a> را دوست دارم.
  </p>
  <p>
    <span rel="foaf:interest" resource="urn:ISBN:0752820907">
      <span about="urn:ISBN:0752820907">
        <cite property="dc:title">
          آینده خود را خلق کنید
        </cite> نوشته
        <span property="dc:creator">برایان تریسی</span> است</span></span>.
      </p>
    </body>
  </html>
```

- ۱- در تگ html با استفاده از صفت prefix نسخه rdf ای که استفاده می کنید را تعیین کنید. (این آدرس در حقیقت آدرس یک Vocabulary است و به موتور جستجو می فهماند که کلماتی که شما برای معنا بخشیدن به عبارات استفاده کرده اید از آن «فرهنگ لغات» یا «فضای نام» انتخاب شده. مثلاً وقتی می خواهید بگویید «ایران»، یک علاقه مندی برای «حمید رضا نیرومند» است، از چه کلمه ای استفاده کرده اید و کجا تعریف شده که این کلمه به معنی علاقه مندی است؟)
- ۲- در تگ body (یا هر تگ کلی دیگر) یک صفت about اضافه کرده و نام فاعل (یا یک آدرس اینترنتی در مورد فاعل) که این صفحه در مورد آن فاعل است را بیان کنید.
- ۳- در صفحه، کلمات مهم مانند نام شخصیت، نام کتاب، .. که با فاعل رابطه ای دارند را با تگی مانند احاطه کنید و به این تگ یک صفت rel (مخفف relation به معنی رابطه) اضافه کنید و در مقدار آن، رابطه ای آن کلمه با فاعل را بنویسید. (مثلاً «کتاب آینده خود را خلق کنید» به عنوان interest یعنی علاقه مندی برای صاحب این صفحه معرفی شده)
- ۴- اگر کلمه ای در صفحه بود که صفت فاعل بود را با تگ span احاطه کنید و یک صفت property به آن اضافه کرده و بگویید که این کلمه چه صفتی برای فاعل است. (مثلاً «حمید رضا نیرومند» یک nickname یا اسم و فامیل برای فاعل به حساب می آید)

نکته: توجه کنید که نام صفات و ارتباطات باید حتماً در vocabulary که در ابتدای صفحه به آن آدرس داده‌اید وجود داشته باشد.

مدل مخفی مارکوف چیست؟

یک مدل آماری است که در آن سیستم مدل شده به صورت یک فرایند مارکوف با State های پنهان فرض می‌شود این مدل برای یافتن State نهایی یک سیستم با توجه به State اولیه کاربرد دارد. (state=حالت) به طور خلاصه:

یعنی ما با توجه به دانسته‌های قبلی و با توجه به یک سری داده اولیه یا حالت آغازین، به یک حالت یا State پایانی برسیم.

مثال:

دو برادر و خواهر به نام‌های آلیس و باب را در نظر بگیرید. آن‌ها دور از هم زندگی کرده و در انتهای هر روز درباره کارهای روزمره‌شان با هم تلفنی صحبت می‌کنند. فعالیت‌های باب شامل "قدم زدن در پارک"، "خرید کردن" و "تمیز کردن آپارتمان" می‌شود. انتخاب اینکه هر روز کدام کار را انجام دهد منحصراً بستگی به هوای همان روز دارد. آلیس اطلاع دقیقی از هوای فعلی محل زندگی باب ندارد ولی از تمایلات کلی وی آگاه است (بنا به نوع هوا چه کاری را دوست دارد انجام دهد). بر اساس گفته‌های باب در پایان روز قبل، آلیس سعی می‌کند هوای آن روز را حدس بزند.

در حقیقت آلیس هوا را یک زنجیره گسسته مارکوف می‌پندارد که دو حالت "بارانی" و "آفتابی" دارد. اما به طور مستقیم هوا را مشاهده نمی‌کند. بنابراین حالات هوا بر او مخفی است. در هر روز احتمال اینکه باب به "قدم زدن"، "خرید کردن" و "تمیز کردن" بپردازد بستگی به هوا داشته و دارای یک احتمال مشخص است و این احتمال را آلیس به مرور و بر اساس صحبت‌های قبلی باب یاد گرفته است. مشاهدات مسأله شرح فعالیت‌هایی است که باب در انتهای هر روز به آلیس می‌گوید.

حالا تصور کنید الان آلیس می‌خواهد به باب زنگ بزند و می‌خواهد بداند او خانه است یا بیرون از خانه؟

مسأله بالا توسط الگوریتم آقای مارکوف به صورت زیر مدل می‌شود:

```
states = ('Rainy', 'Sunny')

observations = ('walk', 'shop', 'clean')

start_probability = {'Rainy': 0.6, 'Sunny': 0.4}

transition_probability = {
    'Rainy': {'Rainy': 0.7, 'Sunny': 0.3},
    'Sunny': {'Rainy': 0.4, 'Sunny': 0.6},
}

emission_probability = {
    'Rainy': {'walk': 0.1, 'shop': 0.4, 'clean': 0.5},
    'Sunny': {'walk': 0.6, 'shop': 0.3, 'clean': 0.1},
}
```

حالات کلی: بارانی یا آفتابی بودن هوا

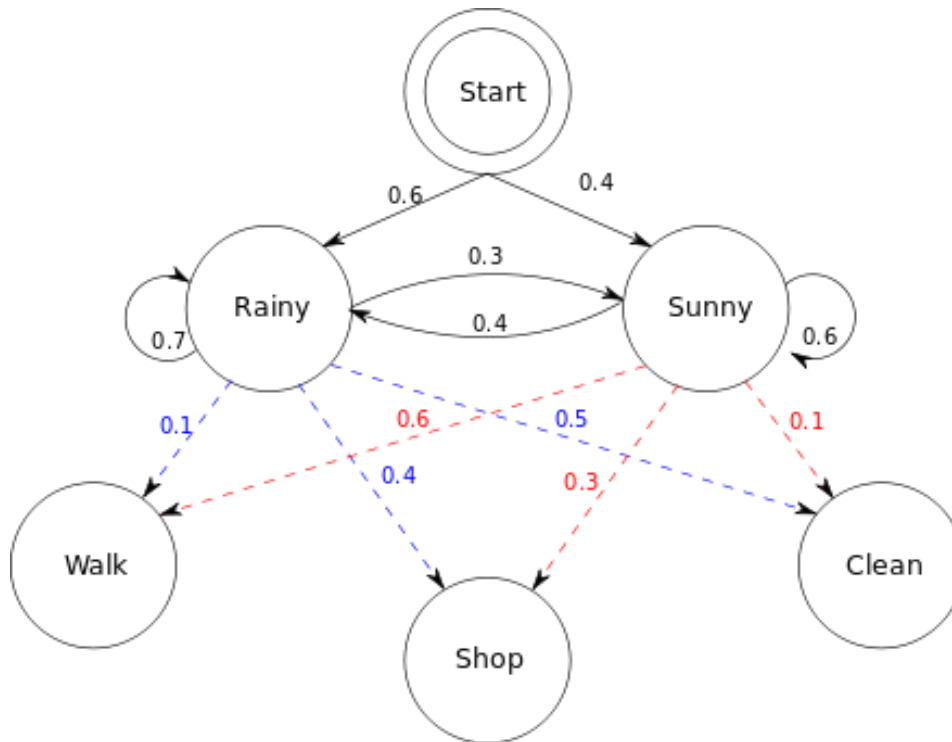
مشاهدات: قدم زدن، خرید کردن، تمیز کردن خانه

احتمالات آغازین (یعنی احتمالاتی که آلیس به مرور و طی مدت‌ها صحبت یاد گرفته): به احتمال ۶۰ درصد هوای محل زندگی باب بارانی است و به احتمال ۴۰ درصد آفتابی.

احتمال انتقال از یک حالت به حالت دیگر: اگر دیشب باب گفته باشد که هوا بارانی است، به احتمال ۷۰ درصد امروز نیز بارانی است و به احتمال ۳۰ درصد آفتابی است و اگر دیشب گفته باشد که هوا آفتابی است، به احتمال ۴۰ درصد امروز آنجا بارانی است و به احتمال ۶۰ درصد همچنان آفتابی است...

احتمال بروز یک کار: اگر بارانی باشد به احتمال ۱۰ درصد باب بیرون قدم می‌زند، به احتمال ۴۰ درصد خرید می‌کند و به احتمال ۵۰ درصد خانه است و خانه را تمیز می‌کند. و اگر هوا آفتابی باشد به احتمال ۶۰ درصد قدم می‌زند، به احتمال ۳۰ درصد خرید است و به احتمال ۱۰ درصد در حال تمیز کردن خانه.

این مدل را می‌شود به صورت گراف زیر نشان داد:



حالا می‌توان راحت‌تر پاسخ داد: آلیس می‌خواهد به باب زنگ بزند، دیشب باب گفته است که هوا آنجا بارانی بوده. به نظر شما الان باب در حال انجام چه کاری است؟

پاسخ: نگاه کنید به حالت یا State یا همان دایره‌ی Rainy، یال‌هایی که عدد بیشتری دارند را دنبال کنید... به احتمال ۷۰ درصد به خودش می‌رود (یعنی وقتی گفته دیشب بارانی بوده به احتمال ۷۰ درصد امروز هم بارانی است) و به احتمال ۵۰ درصد به Clean می‌رود، پس او در حال نظافت و در خانه است...

کاربردهای مدل مخفی مارکوف

این مدل در زمینه‌های بسیار زیادی کاربرد دارد:

- تبدیل گفتار به صدا (تشخیص گفتار):

تصور کنید شما در حال خواندن یک متن برای یک نرم افزار هستید که او آنرا تایپ کند. اگر تلفظ یک

کلمه شبیه یک کلمه دیگر باشد (مثلاً می‌گویید: Important Object نرم افزار ممکن است

an Object بشنود...) چطور تشخیص دهد که کدام منظور شما بوده؟ با مدل مخفی مارکوف و چیزهایی

که قبلاً یادش داده‌اید...

- تشخیص چهره:

مثلاً تصور کنید یک نرم افزار تشخیص چهره، فاصله بین دو مردمک چشم را ملاک شناسایی قرار دهد. اگر برای دو نفر از افراد یک سازمان این فاصله یکسان باشد، کدام یکی الان جلو دستگاه ایستاده؟ با توجه به دانسته‌های دیگر می‌شود فهمید...

- ترجمه ماشینی:

همه با Google Translate کار کرده‌اند و می‌دانند که گاهی چقدر اشتباه ترجمه می‌کند! دلیل؟ هنوز دیتابیس مربوط به مدل مخفی مارکوف گوگل کامل نشده. به همین دلیل است که گوگل، Community Help را راه‌اندازی کرده و از شما انسان‌ها خواسته به مرور به آن نرم‌افزار یاد بدهید که وقتی فلان کلمه کنار فلان کلمه قرار می‌گیرد ترجمه‌اش چه می‌شود؟

- پیش‌بینی ژن

- هم تراز کردن توالی:

تشخیص قرار گرفتن ماشین بر روی خط، تشخیص در یک راستا قرار گرفتن نقاط یک خط

- تشخیص فعالیت:

پزشک هوشمند، خانه‌های هوشمند

نکاتی در مورد موتورهای جستجو

۳ نکته مهم در ساخت یک موتور جستجوی عالی

۱- خزیدن عمیق Crawling deeply

موتور جستجو در لینک‌های موجود بر روی سایت‌های اینترنتی می‌خزد و داده‌های هر صفحه را به دیتابیس خود اضافه می‌کند. هر چه یک موتور جستجو لینک‌های بیشتری را پیمایش کند و این کار را سریع‌تر انجام دهد، داده‌های بیشتری در زمان کمتری به دست خواهد آورد.

۲- نمایه‌سازی Indexing

پس از جمع‌آوری داده‌ها، موتور جستجو باید آن‌ها را نمایه‌سازی کند. هر چه این نمایه‌سازی با معیارهایی که در این کتاب بحث شد، بیشتر منطبق باشد، جستجوی کاربران نتایج مرتبط‌تر و مؤثرتری را در پی خواهد داشت.

۳- رتبه‌دهی Ranking

رتبه‌دهی یکی از ابتکارات مؤسسان گوگل است. آن‌ها به هر صفحه با توجه به معیارهایی که در اصطلاح SEO نامیده می‌شود، یک رتبه بین 0 تا 10 می‌دهند.

SEO

مخفف Search Engine Optimization و به معنی «بهبود در موتورهای جستجو» است. به مجموعه قوانینی که با رعایت آن‌ها در یک سایت، جایگاه و ارزش آن سایت در موتورهای جستجو افزایش پیدا می‌کند گفته می‌شود.

عملیات Document Selection

به عملیات جستجوی تک تک کلمات کلیدی یک عبارت جستجو شده و ساختن اسنادی که این کلمات در متن آن‌ها وجود دارد عملیات "انتخاب سند" گفته می‌شود.

Spider/Crawler

نرم افزارها یا ربات‌هایی که لینک‌های یک سایت را پیمایش می‌کند تا به صفحات مختلف اینترنت برسند و محتوای هر صفحه را در دیتابیس گوگل ذخیره می‌کنند، اصطلاحاً «خزنده» یا «عنکبوت» گفته می‌شود.

Indexer

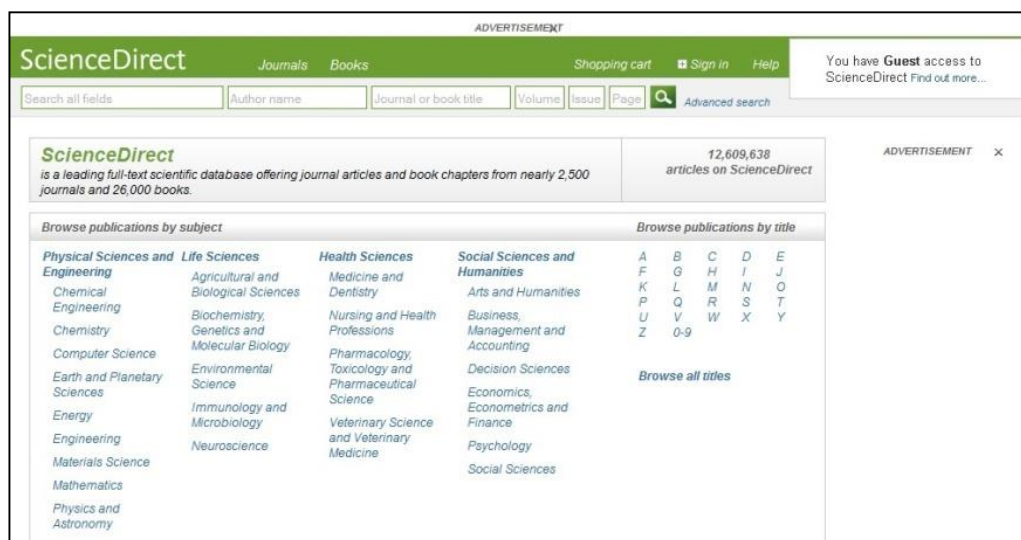
به نرم‌افزاری که اطلاعاتی که توسط خزنده جمع‌آوری شده است را نمایه‌زنی می‌کند، ایندکسر گفته می‌شود.

معرفی چند سایت پژوهشی ایرانی و خارجی

سایت‌های خارجی

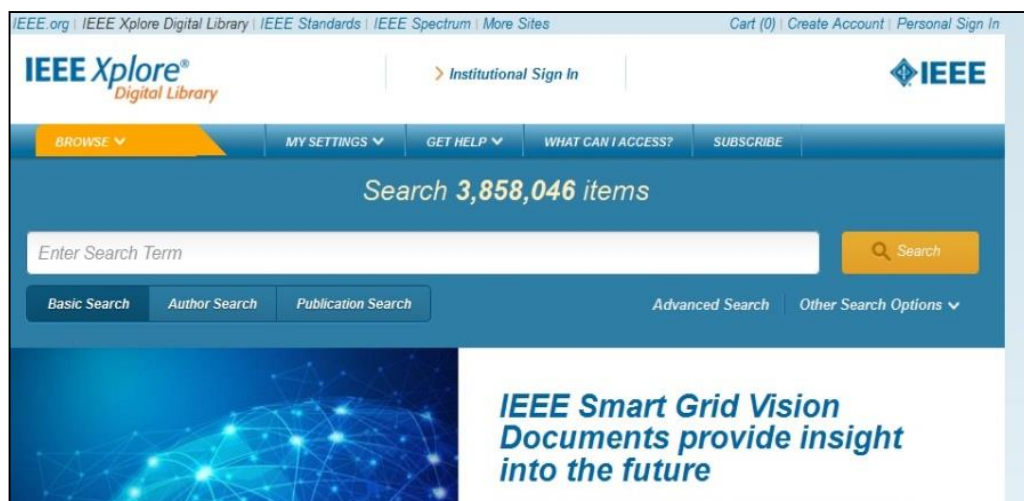
1- sciencedirect.com

این سایت مهم‌تری مرجع محققان دانشگاهی است. دسته‌بندی الفبایی و موضوعی دارد و تا این لحظه بیش از ۲۵۰ هزار مجله و ۲۶۰ هزار کتاب در آن موجود است.



2- ieee.org

ویژه مقالات پژوهشی مرتبط با کامپیوتر، الکترونیک و مخابرات است. داغ‌ترین مطالب روز را می‌توان در قسمت most popular پیدا کرد.



3- ISI.edu

در زمینه پزشکی و علمی کاربرد دارد و از لحاظ اعتبار اولین اعتبار را در مقالات دارد.



سایت‌های ایرانی

4- civilica.com

محلی برای اطلاع از آخرین همایش‌های پژوهشی و همینطور متن مقالات ارائه شده در آن‌ها. انواع ارائه‌ی مقاله در کنفرانس‌ها:

- ۱- ارائه حضوری برای حضار در همایش (بهترین حالت)
- ۲- ارائه حضوری برای متخصصان همایش
- ۳- چاپ مقاله در مجله مربوط به کنفرانس
- ۴- چاپ پوستر (در این حالت یک چکیده از مقاله شما و مشخصات شما در راهروهای نمایشگاه به نمایش گذاشته می‌شود)

5- sid.ir

پایگاه داده اطلاعات علمی که توسط جهاد دانشگاهی تأسیس شده است و حاوی بانک طرح‌های پژوهشی و بانک نشریات ایران است. اولین پایگاه علمی ایران با دسترسی آزاد (Open Access) به حساب می‌آید.



6- irandoc.ac.ir

۷- حاوی چکیده پایان‌نامه‌های دانشجویان ایرانی.

اطلاعه‌ها	رویدادها	تازه‌ها	اخبار	خدمات و محصولات پژوهشگاه
<ul style="list-style-type: none"> برگزاری ۲۷ دوره‌ی آموزشی پژوهشگاه در آبان ۹۳ انتشار کتاب سنجش و مصورسازی پراکندگی جغرافیایی ثروت علمی؛ مفاهیم و کاربردها برگزاری میزگرد تخصصی بردارش زبان طبیعی و علوم شناختی آغاز سال تحصیلی جدید بر جامعه علمی کشور و پویندگان دانش مبارک باد. اطلاعیه ثبت‌نام پذیرفته‌شدگان آزمون نیمه‌متمرکز دکتری، به‌توقف به‌سر سال ۱۳۹۳، پژوهشگاه 	<ul style="list-style-type: none"> تقدیر معاون آموزشی وزارت علوم از ایراندک برگزاری کارگاه تحلیل شبکه اجتماعی برای اولین بار در ایراندک تحلیل از برگزیدگان حوزه نشر الکترونیک با حضور معاون وزیر ارشاد در ایراندک سامانه پیشنهادی پژوهش در ایراندک آغاز به کار کرد شماره جدید «پژوهشنامه بردارش و مدیریت اطلاعات» 			<ul style="list-style-type: none"> نرم‌افزار ساخت اصطلاحنامه طراحی، تولید و تدوین اصطلاحنامه شبکه اجتماعی دیجیتال‌سازی و ساخت آرشیو دیجیتال برنامه‌ریزی راهبردی سامانه مدیریت یکپارچه نشریات رایاسمین