

پیامبر خدا (ص):
 حُسْنُ الْمَسْأَلَةِ نِصْفُ الْعِلْمِ.
 خوب پرسیدن، نیمی از دانش است.
 تحف العقول : ۵۶ منتخب میزان الحکمة : ۲۶۰

Business Intelligence

درس: هوشمندی کسب و کار (هوش تجاری)
 بخش دوم: آشنایی با تکنیک ها و تکنولوژی های هوش تجاری
 فصل نهم: متن کاوی و وب کاوی

گروه: مهندسی فناوری اطلاعات دانشگاه نور طوبی - رشته: تجارت الکترونیک
 مدرس: حسنعلی نعمتی شمس آباد
 دستیار (TA): زهرا سادات طبائی عقدائی

هدف فصل نهم: متن کاوی و وب کاوی

- هدف از این فصل آشنایی با متن کاوی و وب کاوی است.
- کاوش در داده های بدون ساختار متنی در این فصل مورد توجه است
- موضوعات زیر در این فصل بررسی می شود:
 - مقایسه متن کاوی با وب کاوی
 - کاربردها و ابزارهای متن کاوی
 - مباحث مقدماتی پردازش زبان طبیعی (NLP) و کاربرد آن در متن کاوی
 - فرآیند سه مرحله ای متن کاوی
 - مباحثی در باره وب کاوی

مفاهیم متن کاوی

3

nemati@ut.ac.ir

حدود ۸۵ تا ۹۰ درصد از داده های هر سازمان، داده های بدون ساختار هستند.

داده های بدون ساختار سازمان، هر ۱۸ ماه حدوداً دوبرابر می شود.

بهره برداری مناسب از چنین منابع اطلاعاتی و تصمیم گیری هوشمندانه بر اساس آنها، یک نیاز رقابتی برای سازمان ها محسوب می شود.

پاسخ به این نیاز:
متن کاوی

✓ فرآیند نیمه خودکار استخراج دانش از منابع داده بدون ساختار
✓ کاوش در داده های متنی یا کشف دانش از پایگاه داده های متنی

متن کاوی در مقایسه با داده کاوی

nemati@ut.ac.ir

تفاوت

داده کاوی داده های ساختار یافته را مورد کاوش قرار می دهد اما متن کاوی داده های بی ساختار را بررسی می کند.
ساختار یافته: پایگاه داده ها.
بی ساختار: متون Word، فایل های PDF، فایل های XML و ...

شابهت

هر دو به دنبال یافتن الگوهای جدید و مفید می باشند.
هر دو فرآیند های نیمه خودکار هستند.

متن کاوی ابتدا خودش ساختاری را به داده ها تحمیل می کند، سپس به کاوش داده های ساختاریافته می پردازد.

متن کاوی = ساختار دهی + داده کاوی

متن کاوی = ساختار دهی + داده کاوی

5

nemati@ut.ac.ir

در متن کاوی برای ساختار دهی به داده های بدون ساختار، ابتدا شاخص های عددی معناداری را از متون بدون ساختار استخراج می کنیم و سپس این شاخص ها را با استفاده از الگوریتم های داده کاوی پردازش می کنیم.

با استفاده از این شاخص های عددی می توان وظایف زیر را انجام داد:

تعیین روابط بین
مستندات و استخراج
محتوای پنهان
مستندات

خوشه بندی مستندات
بر اساس مفاهیم مشابه
و موضوعات مشترک

جمع بندی و خلاصه
سازی مستندات بر
اساس مفاهیم کلیدی

برخی از حوزه های کاربردی متن کاوی (۱)

nemati@ut.ac.ir

ارتباطات الکترونیکی (مثلا Email)

- فیلتر هرزنامه (Spam)
- طبقه بندی و اولویت بندی ایمیل ها
- تولید پاسخ خودکار

داده های غنی متن (text-rich data)

- قانونی (دستورات دادگاه)
- تحقیقات دانشگاهی (مقالات پژوهشی)
- مالی (گزارشات ماهانه)
- پزشکی (پرونده های پزشکی)
- زیست شناسی (واکنش های مولکولی)
- تکنولوژی (فایل های ثبت اختراعات)
- بازاریابی (نظرات مشتریان)

برخی از حوزه های کاربردی متن کاوی (۲)

nemati@ut.ac.ir

بازاریابی

- تجزیه و تحلیل صداها و متون نظرات، پیشنهادات، انتقادات و شکایات مشتریان و در نتیجه بهبود شرایط و افزایش خرید مشتریان.
- تحلیل متون بی ساختار و ترکیب آن با داده های ساختاریافته سازمانی جهت پیش بینی خواسته ها و رفتار خرید مشتریان.



امنیت

- تعیین خودکار محتوای تماس های تلفنی، فکس ها، ایمیل ها، اطلاعات ماهواره ای و سایر شبکه های تلفنی جهت شناسایی جرم ها.
- تشخیص فریبکاری ها و کلاه برداری ها.



برخی از حوزه های کاربردی متن کاوی (۳)

nemati@ut.ac.ir

زیست پزشکی

- تحلیل های مربوط به DNA و شناسایی ژن ها.
- تعیین محل پروتئین درون سلولی جهت اهداف دارویی.
- استخراج ارتباطات ژن - بیماری.



متون دانشگاهی

- تجزیه و تحلیل جریان های پژوهشی.
- جستجوی مفهومی متون مرتبط .



کاربردهای متن کاوی (۱)

- مشخص کردن عبارت های کلیدی و تعیین رابطه بین آنها بر اساس ترتیب به کار رفته در متن با توجه به تطابق الگوها - پردازش زبان طبیعی برای تحلیل و پردازش متون (مثلا خواندن هزاران رزومه و استخراج اطلاعات آنها)

Information extraction

استخراج
اطلاعات

- پیش بینی اسناد مورد علاقه کاربران بر اساس پروفایل کاربر و اسنادی که قبلا مطالعه کرده است.

Topic tracking

ردیابی
موضوع

- خلاصه سازی خودکار اطلاعات جهت صرفه جویی در وقت کاربران. برای مثال جمع آوری اطلاعات از سایت های خبری.

Summarization

خلاصه
سازی

کاربردهای متن کاوی (۲)

nemati@ut.ac.ir

- تعیین موضوعات اصلی (تم های اصلی) یک سند و قرار دادن آن در طبقه از پیش تعیین شده براساس موضوعات.

Categorization

رده بندی

- گروه بندی اسناد مشابه بدون داشتن مجموعه طبقات از پیش تعیین شده.

Clustering

خوشه بندی

- ایجاد ارتباط بین اسناد مرتبط بر اساس مفاهیم مشترک و بهبود روش های جستجوی سنتی

Concept linking

ایجاد رابطه بین مفاهیم

- یافتن بهترین پاسخ برای سؤال ارائه شده از طریق انطباق الگوهای دانش گرا (Knowledge-Driven Pattern Matching).

Question answering

پاسخ گویی به پرسش ها

تمرین ۹-۱

آیا متن کاوی همان جستجو بر روی متن نیست؟ چرا؟
در مورد مفهوم واژگان زیر در متن کاوی تحقیق نمایید.

- Corpus (and corpora)
- Terms & Term dictionary
- Concepts
- Stemming
- Stop words (and include words)
- Synonyms (and polysemes)
- Tokenizing
- Word frequency
- Part-of-speech tagging
- Morphology
- Term-by-document matrix
 - Occurrence matrix
- Singular value decomposition
 - Latent semantic indexing

ساختار دهی به متون

nemati@ut.ac.ir



دیدگاه های ساختاردهی به متون



دیدگاه جدید

Natural Language Processing

دیدگاه قدیمی

Bag-Of- Word

پردازش زبان طبیعی (NLP):

- در متن کاوی اهمیت فراوانی دارد.
- زیرشاخه از هوش مصنوعی و زبان شناسی محاسباتی است.
- هدف درک معنای در زبان های طبیعی است.

کشکول کلمه (BOW):

- تبدیل متن به مجموعه ای از لغات، بدون توجه به گرامر و یا ترتیب قرار گرفتن لغات.
- قرار گرفتن متون در دسته های ساده از پیش تعیین شده و یا گروه بندی آنها به چندین گروه متمایز.

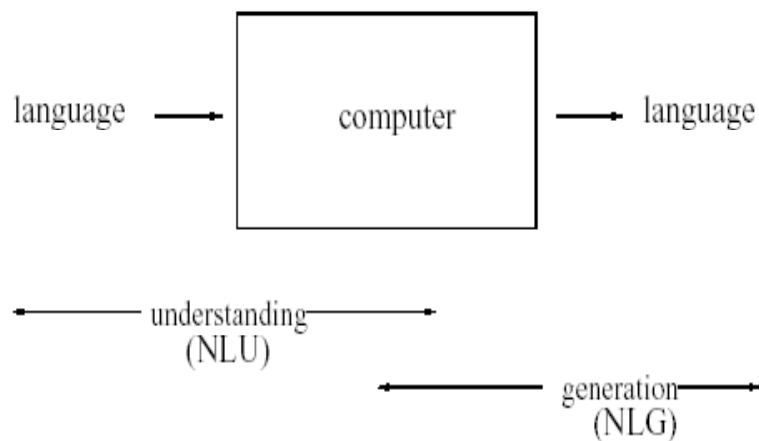
پردازش زبان طبیعی (NLP)

Natural Language Processing

nemati@ut.ac.ir

- موضوع مطالعه NLP درک زبان طبیعی انسان توسط ماشین است و یکی از اجزای اصلی متن کاوی به حساب می آید.

پردازش زبان طبیعی عبارت است از استفاده از کامپیوتر برای پردازش زبان گفتاری و نوشتاری انسان.



هدف نهایی از پردازش زبان طبیعی این است که کامپیوترها از زبان طبیعی به عنوان ورودی و خروجی استفاده کنند.

موانع اصلی و چالش های پیاده سازی NLP

- زبان های طبیعی بسیار متنوع و پیچیده هستند و دو مانع اصلی برای درک زبان طبیعی توسط ماشین وجود دارد:
- درک معانی بسیاری از کلمات در یک جمله معمولا مستقل از گرامر است.
- گرامر بسیاری از زبان های طبیعی دقیق نیست.
- با توجه به موانع اساسی فوق چالش های فراوانی برای پیاده سازی NLP وجود دارد.

چالش های موجود در پیاده سازی NLP (۱)

nemati@ut.ac.ir

- برچسب گذاری عبارتهای موجود در یک سند و تشخیص اینکه اسم، صفت یا ضمیر است دشوار است. چرا که تنها به تعریف ادات (اسم، صفت، فعل و ...) بستگی ندارد بلکه به زمینه ای که ادات در آن به کار رفته وابسته است.

Part- of speech tagging

برچسب
گذاری ادات
سخن

- در برخی زبان های طبیعی همچون فارسی، چینی و ژاپنی یک واژه دارای مرزهای معنایی گوناگونی است و تجزیه متن (Text-Parsing) نیازمند شناسایی مرزهای معنایی کلمه است.

Text segmentation

بخش بندی
متن ها

- برخی از کلمات دارای معانی متفاوت و گوناگونی هستند که انتخاب معنای مناسب برای آن دشوار است.

Word sense disambiguation

ابهام زدایی
معنایی کلمه

چالش های موجود در پیاده سازی NLP (۲)

nemati@ut.ac.ir

- گرامر یک زبان طبیعی اغلب مبهم است. انتخاب مناسب ترین ساختار نیازمند ادغام اطلاعات متنی و معنایی است.

Syntactic ambiguity

ابهام نحوی

- لهجه های خارجی یا منطقه ای و یا موانع صوتی در سخنرانی و خطاهای چاپی یا دستوری در متون، پردازش زبان را دشوار می سازند.

Imperfect or irregular input

ورودی های ناقص یا غیر معمول

- بیان یک جمله توسط گوینده ممکن است با حرکاتی معنادار و بیانی خاص همراه باشد. بنابراین ساختار جمله به تنهایی همه اطلاعات موجود در آن را شامل نمی شود.

Speech acts

کنش های کلامی

کاربردها و وظایف NLP (۱)

nemati@ut.ac.ir

- علم جستجوی اسناد مرتبط، یافتن اطلاعات مخصوص آنها و تولید فراداده های مربوط به محتوای آنها. **Information retrieval**

بازیابی
اطلاعات

- استخراج خودکار اطلاعات ساختاریافته. برای مثال طبقه بندی متون بر اساس دامنه موضوعی آنها. **Information extraction**

استخراج
اطلاعات

- یکی از زیر وظایف استخراج اطلاعات می باشد که عناصر اتمیک را در متن شناسایی نموده و در طبقه های از پیش تعیین شده قرار می دهد. **Named-entity recognition**

تشخیص
نامی -
موجودیتی

- جهت ایجاد پاسخ زبان انسانی، یک برنامه کامپیوتری از پایگاه داده های از پیش ساختاریافته و مجموعه ای از اسناد زبان انسانی، استفاده می کند. **Question answering**

پاسخ گویی
به سؤالات

کاربردها و وظایف NLP (۲)

nemati@ut.ac.ir

- ایجاد مجموعه ای خلاصه شده از اسناد متنی که شامل نقاط مهم متن اصلی می باشد.
- Automatic summarization**

خلاصه سازی
خود کار

- سیستم هایی اطلاعات را از پایگاه داده های کامپیوتری به زبان بشری خوانا تبدیل می کنند.
- Natural Language Generation-NLG**

ایجاد زبان
طبیعی

- سیستم هایی نمونه هایی از زبان انسانی را به قالب های رسمی برای کامپیوتر که راحت تر قابل بررسی هستند تبدیل می کنند.
- Natural Language Understanding - NLU**

درک زبان
طبیعی

- ترجمه خود کار از یک زبان به زبان انسانی دیگر.
- Machine translation**

ترجمه ماشینی

- ترجمه خود کار تصویری از دست خط، متن تایپی یا متن چاپی به اسناد متنی قابل ویرایش ماشینی.
- Optical character recognition**

تشخیص
کاراکتر نوری

کاربردها و وظایف NLP (۳)

nemati@ut.ac.ir

- یک برنامه کامپیوتری که به یک سخنران غیر بومی کمک می کند تا متون زبان خارجی را با لهجه و تلفظ صحیح لغات بخواند.

خواندن زبان
خارجی

- یک برنامه کامپیوتری که به کاربر غیر بومی در نوشتن یک متن خارجی کمک می کند.

Foreign language reading & writing

نوشتن زبان
خارجی

- تبدیل واژه های گفتاری به ورودی های قابل خواندن توسط ماشین و تبدیل آن به متن.

Speech recognition

تشخیص
گفتار

- یک برنامه کامپیوتری که به صورت اتوماتیک متن زبان معمولی را به گفتار تبدیل می کند.

Text-to-Speech or Speech Synthesis

تبدیل متن به
گفتار

- یک برنامه کامپیوتری که متنی را خوانده و خطاهای آن را شناسایی و تصحیح می کند.

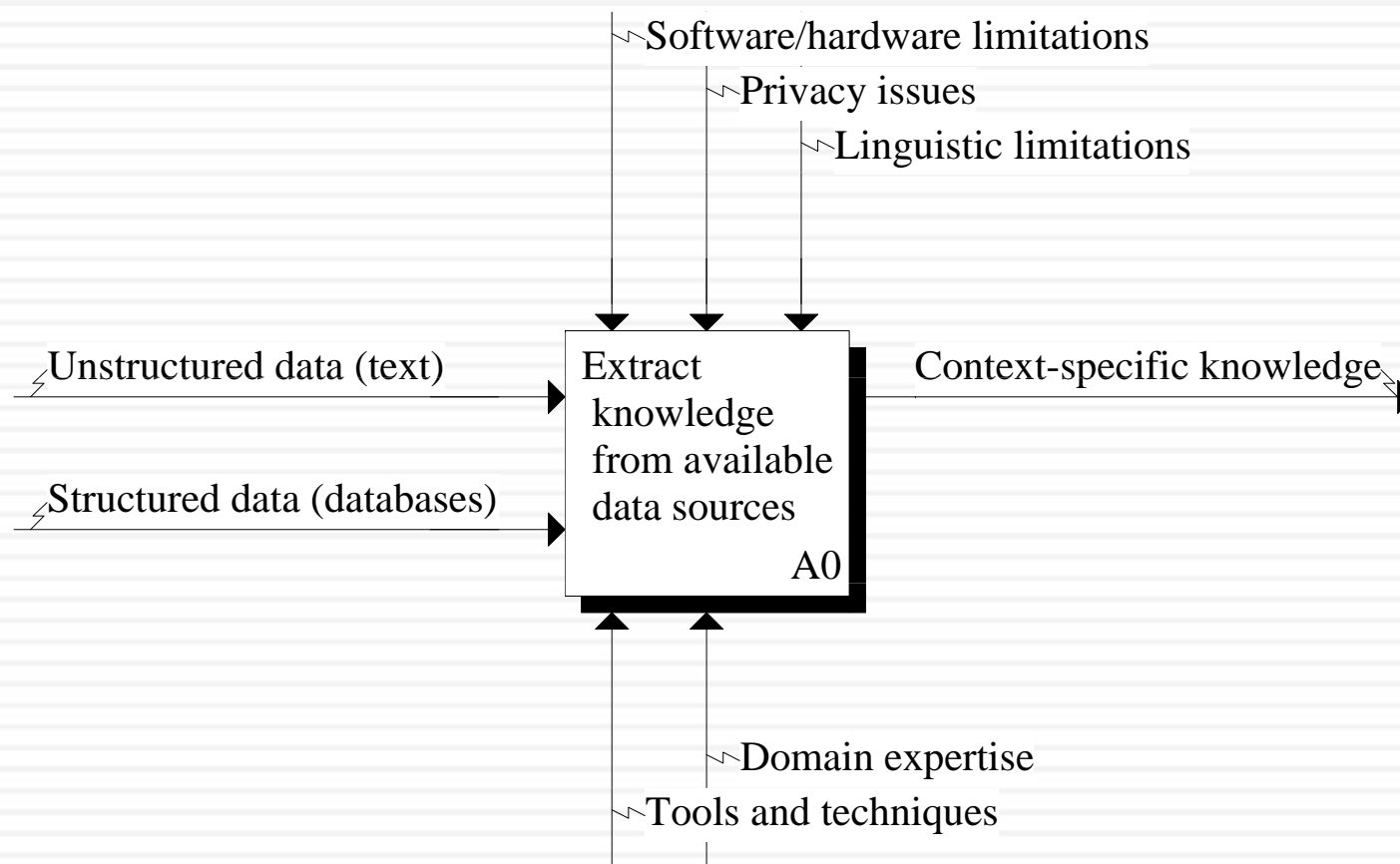
Text proofing

تصحیح متن

فرآیند متن کاوی (۱)

nemati@ut.ac.ir

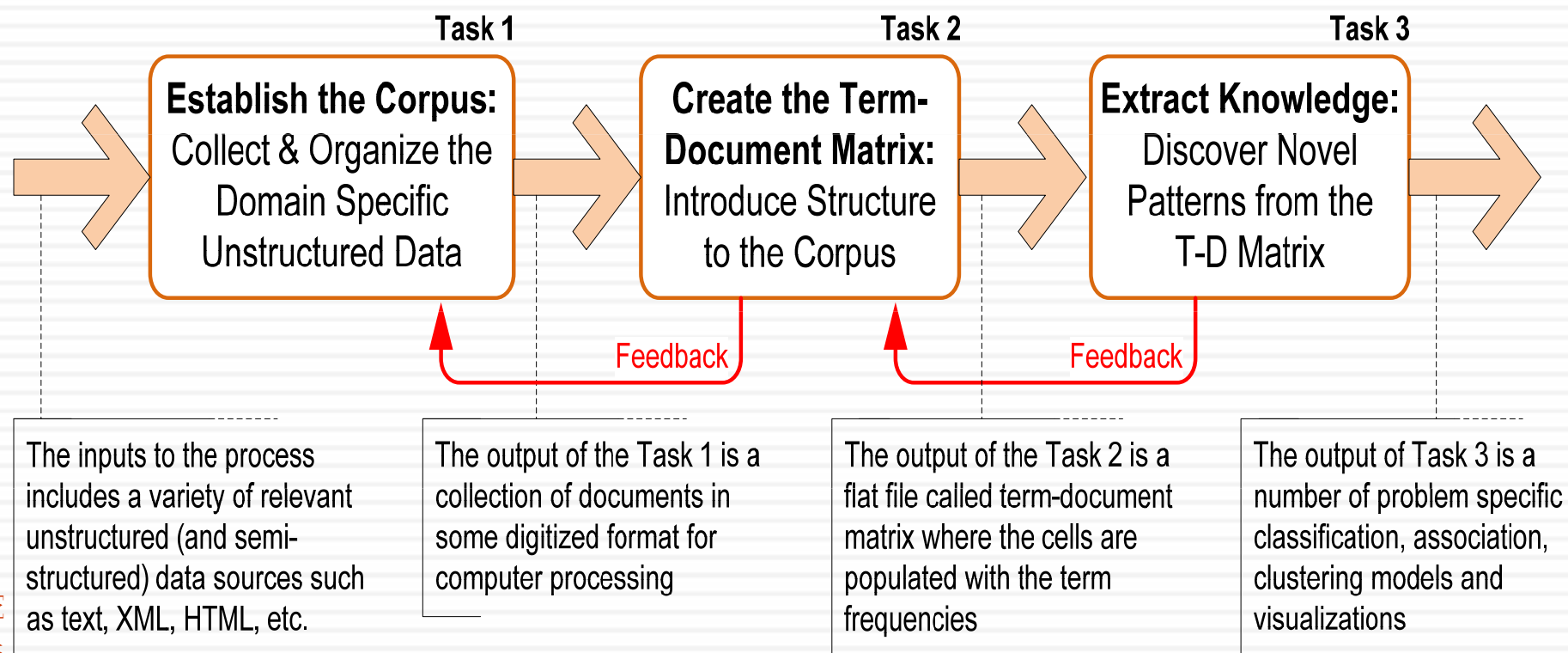
□ نمودار اولیه جریان داده ها در فرآیند داده کاوی



نمودار زمینه ای (Context Diagram) یا DFD سطح صفر

فرآیند متن کاوی (۲)

nemati@ut.ac.ir



منبع مورد استفاده: [۱]

مرحله ۱: ایجاد مجموعه متن (Corpus)

nemati@ut.ac.ir

جمع آوری مجموعه داده های بدون ساختار (متن ها) مرتبط با هم در قالب های (صفحات وب، اسناد متنی، فایل های XML، ایمیل ها، یادداشت های کوتاه، صداهای ضبط شده)

دیجیتال سازی و استانداردسازی مجموعه متن.
(برای مثال تبدیل فرمت به فایل های متنی ASCII)

قراردادن مجموعه متن در یک مکان مشخص.
(برای مثال در یک فایل یکپارچه و یا در یک دایرکتوری و یا در فایل های جداگانه)

مرحله ۲: ایجاد ماتریس عبارت در متن (TDM)

(۱) Term-by-Document Matrix

nemati@ut.ac.ir

Terms \ Documents	investment risk	project management	software engineering	development	SAP	...
Document 1	1			1		
Document 2		1				
Document 3			3		1	
Document 4		1				
Document 5			2	1		
Document 6	1			1		
...						

مرحله ۲: ایجاد ماتریس عبارت در متن (TDM)

(۲)

nemati@ut.ac.ir

گام اول: حذف واژگان عمومی

- حذف کلمات و حروف اضافی (مانند "و"، "یک"، "از" و ...) که کلمات وقف (Stop Word) نامیده می شوند.

گام دوم: ریشه یابی

- جایگزینی واژگان با ریشه های آنها.
- علامت جمع، علائم ربط یا صرف حذف می شوند. برای مثال "تلفن کردن"، "تلفن شده" و "تلفن ها" به صورت "تلفن" نوشته می شوند.

مرحله ۲: ایجاد ماتریس عبارت در متن (TDM)

(۳)

nemati@ut.ac.ir

گام سوم: توجه به کلمات مترادف

- واژگان مترادف به یک صورت آورده می شوند. برای مثال شاگرد و دانش آموز در یک گروه قرار می گیرند.
- شناسایی عبارات نیز اهمیت دارد. برای مثال منظور از **Microsoft Windows** سیستم عامل است اما منظور از **Window** پنجره است.

گام چهارم: تعیین وزن واژگان

- عامل تکرار: تعداد دفعاتی که یک واژه در یک متن تکرار شده است (فراوانی).
- عامل تکرار معکوس: تعداد دفعاتی که یک واژه در کل متون تکرار شده است.

سوالات اساسی برای ساخت TDM (۱)

- بهترین شاخص جهت تعیین وزن عبارت مورد نظر چیست؟
- چگونه می توان ابعاد ماتریس را به اندازه مناسبی کاهش داد؟

سؤالات اساسی برای ساخت TDM (۲)

nemati@ut.ac.ir

□ بهترین شاخص جهت تعیین وزن عبارت مورد نظر چیست؟

تابع تبدیل لگاریتمی با متغیر wf

$$f(wf) = 1 + \log(wf) \quad \text{for } wf > 0$$

فراوانی واژه در یک سند

فراوانی لگاریتمی
Log Frequencies

$$f(wf) = 1 \quad \text{for } wf > 0$$

فراوانی دودویی
Binary Frequencies

سوالات اساسی برای ساخت TDM (۳)

nemati@ut.ac.ir

□ بهترین شاخص جهت تعیین وزن عبارت مورد نظر چیست؟

$$idf(i, j) = \begin{cases} 0 & \text{if } wf_{ij} = 0 \\ (1 + \log(wf_{ij})) \log \frac{N}{df_i} & \text{if } wf_{ij} \geq 1 \end{cases}$$

تعداد اسناد

فرایوانی واژه i در کل اسناد

فرایوانی عبارت i در سند j

معکوس فرایوانی سند
Inverse Document Frequencies

سوالات اساسی برای ساخت TDM (۴)

nemati@ut.ac.ir

□ چگونه می توان ابعاد ماتریس را به اندازه مناسبی کاهش داد؟

خبره با مشاهده لیست به صورت دستی بخش های غیر ضروری را حذف نماید.

حذف واژه هایی که بسیار کم در تعداد اندکی از اسناد رخ می دهند.

تبدیل ماتریس با استفاده از تجزیه ارزش منفرد (**Singular Value Decomposition**):
کل ابعاد ماتریس ورودی را کاهش می دهد به گونه ای که ابعاد متوالی بیشترین
درجه تنوع را (بین واژه ها و اسناد) دارند.

مرحله ۳: استخراج دانش

nemati@ut.ac.ir

- قرار دادن متون در کلاس های از قبل تعیین شده بر اساس موضوع، عنوان و مفهوم.

رده بندی
(Classification)

- گروه بندی اسناد بر اساس تشابه بین آنها در گروه های گوناگون.
- بهبود دقت و افزایش صحت جستجوی کاربر.

خوشه بندی
(Clustering)

- ایجاد ارتباط معنادار بین عبارات یا مجموعه ای از مفاهیم.

قواعد وابستگی
(Association rules)

- نوع و گرایش جمع آوری داده در توزیع مفاهیم تأثیر دارد.

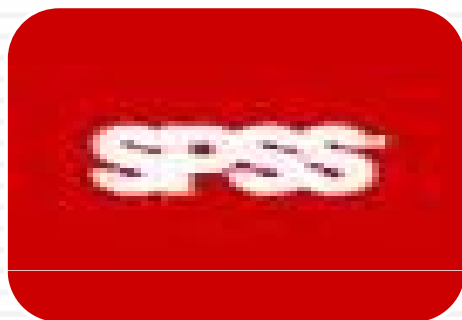
تحلیل روند
(Trend Analysis)

ابزارهای متن کاوی

- با توجه به درک ارزش متن کاوی در سازمان، ابزارهای نرم افزاری متنوعی برای متن کاوی تولید شده است که می توان آنها را به دو دسته تقسیم بندی کرد:
 - ابزارهای تجاری متن کاوی
 - ابزارهای رایگان متن کاوی
- هر سیستم نرم افزاری متن کاوی سه بخش اصلی دارد:
 - مدیریت مستندات
 - پردازش مستندات
 - ابزارهای داده کاوی

نمونه ابزارهای تجاری متن کاوی

nemati@ut.ac.ir



Spss Lexiquista
Spss.com



متن کاو هوشمند IBM
Ibm.com



متن کاو SAS
Sas.com



داده کاو StatSoft
Statsoft.com



تحلیل گر متن هوشمند
megaputer.com



متن کاو خبره
inightful.com

نمونه ابزارهای رایگان متن کاوی

nemati@ut.ac.ir

GATE جعبه ابزار متن باز □

Gate.ac.uk □

RapidMiner □

Rapid-i.com □

LingPine □

Alia-i.com/lingpine □

S-EM(Spy-EM) □

Cs.uic.edu □

Vivsimo/Clusty □

Clusty.com □

اجزای اصلی سیستم های متن کاوی

nemati@ut.ac.ir

بخش مدیریت مستندات اشکال مختلف (متن ساده، پردازش متنی، PDF) را از منابع مختلف (مثل فایل ها، فرم های شبکه ای و نامه های الکترونیکی) مدیریت می کند.

بخش پردازش مستندات که از مستندات فوق استفاده کرده و فایل های داده ای جهت کاوش را تولید می کنند. در قسمت پردازش مستندات کارهای مانند بخش بخش کردن جملات، تعیین کلمات، علامت زنی متن، تجزیه متن و خلاصه سازی متن انجام می شود.

ابزارهای داده کاوی مثل الگوریتم های خوشه بندی، رده بندی و ...

تمرین ۹-۲

- با مراجعه به مقاله های متن کاوی، جزئیات فرآیند سه مرحله ای متن کاوی را بررسی کنید.
- از چه تکنیک ها و ابزارهایی در فرآیند متن کاوی منظور استفاده شده است؟
- هر یک از گام های تولید مجموعه متن (Corpus) چگونه برداشته شده است؟
- چه روشهایی برای ایجاد ماتریس TDM به کار گرفته شده است؟ چه روشهایی برای تعیین وزن واژه ها و کاهش ابعاد ماتریس TDM وجود دارد؟ مقاله ای را پیدا کنید که در آن از روش تجزیه ارزش منفرد (Singular Value Decomposition - SVD) و یا روشهای تحلیل مؤلفه اصلی (Principle Component analysis) استفاده شده است.
- در باره دو روش متداول زیر که برای خوشه بندی متن ها در مرحله سوم استفاده می شود تحقیق کنید:
 - Scatter/Gather Clustering
 - Query-Specific Clustering

□ ژ

مفهوم وب کاوی

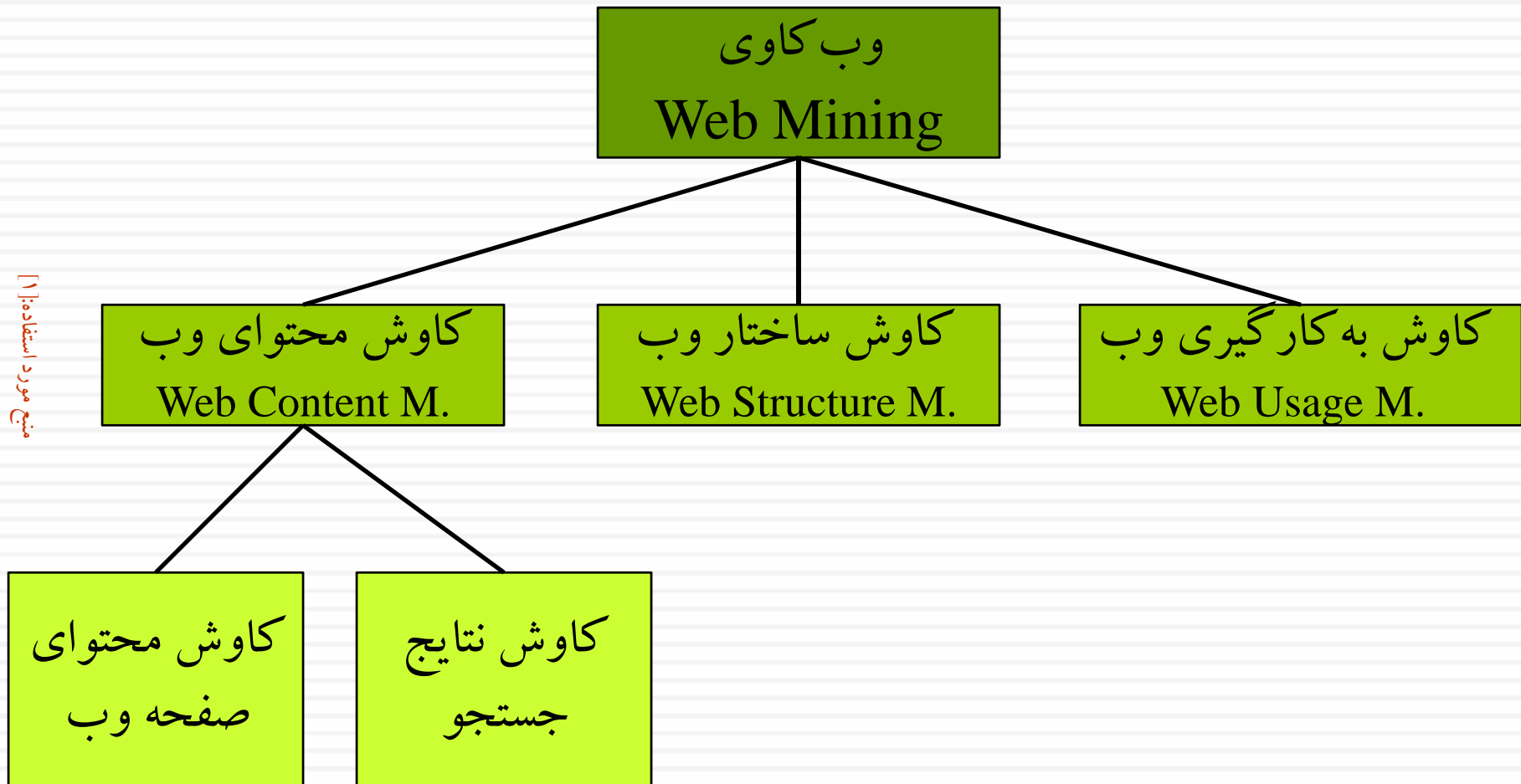
nemati@ut.ac.ir

- وب بزرگترین مخزن داده/متن در دنیا است که حجم اطلاعات آن هر روزه در حال افزایش است.
- وب پیچیده و پویاست؛ بازدیدکنندگان آن دارای علایق، پیش زمینه و نیازمندی های متفاوتی هستند؛ بسیاری از اطلاعات وب مفید نیستند. هر بازدیدکننده از یک وبسایت، هر جستجو از یک موتور جستجو، هر کلیک بر یک لینک و هر مبادله در یک سایت تجارت الکترونیکی به ایجاد داده های جدیدی منجر می شود.

**وب کاوی: کشف و تحلیل اطلاعات مورد علاقه و مفید از وب،
درباره وب و معمولاً از طریق ابزارهای متنی وب.**

انواع وب کاوی

nemati@ut.ac.ir



منبع مورد استفاده: [۱]

منبع: محتوای صفحات

منبع: URL موجود در صفحات

منبع: فایل جزئیات بازدید سایت

کاوش محتوای وب

nemati@ut.ac.ir

به استخراج اطلاعات مفید از صفحات وب اشاره می کند.

مستندات را به صورت قابل خواندن توسط ماشین استخراج می کند.

می توان جهت خواندن محتوای وب به صورت خودکار از نرم افزارهای پیمایش وب استفاده نمود.

می توان از تحلیل محتوای وب برای ارتقای نتایج حاصل از موتورهای جستجو بهره گرفت.

کاوش ساختار وب

nemati@ut.ac.ir

به استخراج اطلاعات مفید از لینک های موجود در مستندات وب اشاره می کند.

اطلاعات استخراج شده برای تعیین میزان عمومیت مستندات و به عنوان مبنایی در تعیین الگوریتم رتبه بندی صفحات وب که اساس موتور جستجوی گوگل است، قابل استفاده هستند.

لینک های موجود می توانند عمق پوشش موضوع را نشان دهند.

شناخت خوب ساختار یک وب سایت برای تعیین اینکه چه صفحاتی باید به مجموعه وبسایت اضافه شوند کمک می کند.

تحلیل لینک ها برای شناخت روابط متقابل، شناخت بهتر یک جامعه، انجمن و یا گروه شبکه ای خاص مفید است.

کاوش به کارگیری وب

nemati@ut.ac.ir

به استخراج اطلاعات مفید از داده های حاصل از بازدید از صفحات وب، مبادلات و سایر موارد اشاره می کند.

داده های حاصل از بازدید وب



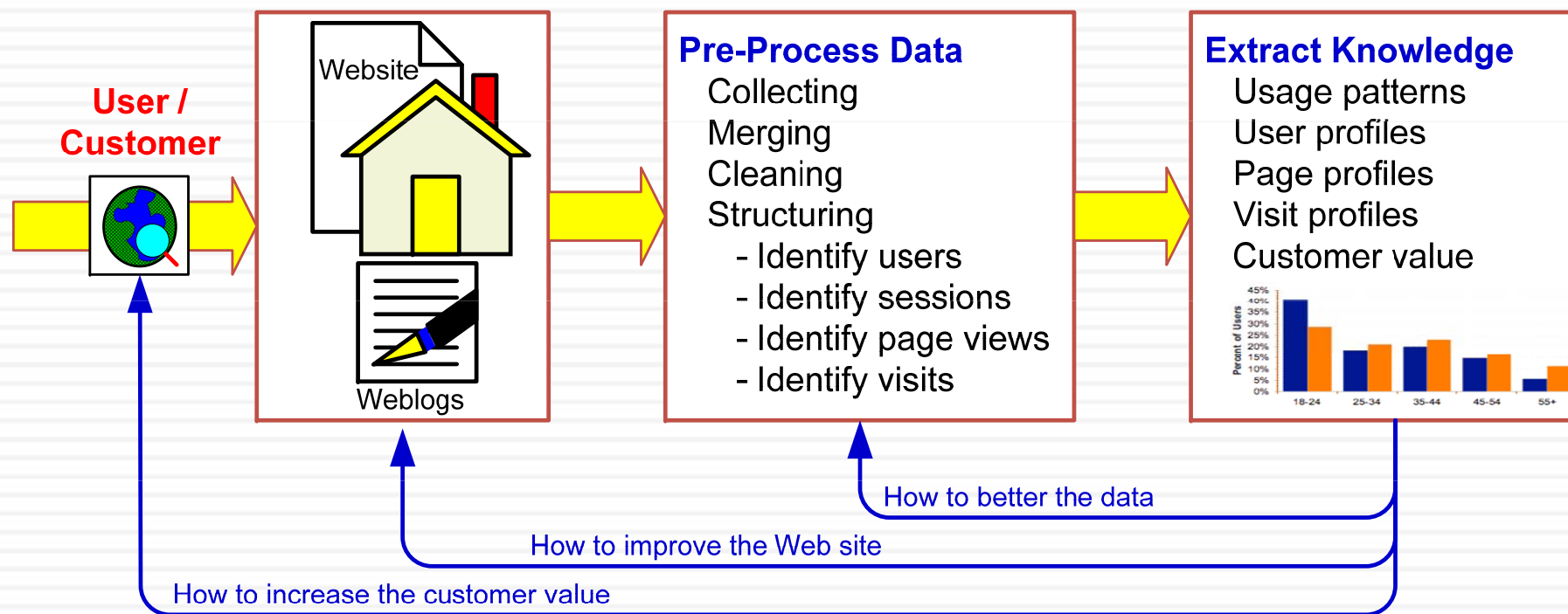
داده هایی که به صورت خودکار ایجاد شده و در مخزن دسترسی سرور، مخزن ارجاع دهنده، مخزن سیستم عامل و کوکی های کاربر ذخیره می شوند.

پروفایل کاربر.

فراداده هایی مثل خصوصیات صفحه، خصوصیات محتوا و به کارگیری داده ها.

استخراج دانش از داده های مربوط به استفاده از وب

nemati@ut.ac.ir



کاربردهای کاوش به کارگیری وب

nemati@ut.ac.ir

طراحی راه کارهای بازاریابی
مشترک برای محصولات

تعیین ارزش طول عمر مشتری

هدف گذاری تبلیغاتی متناسب با
الگوی بازید افراد

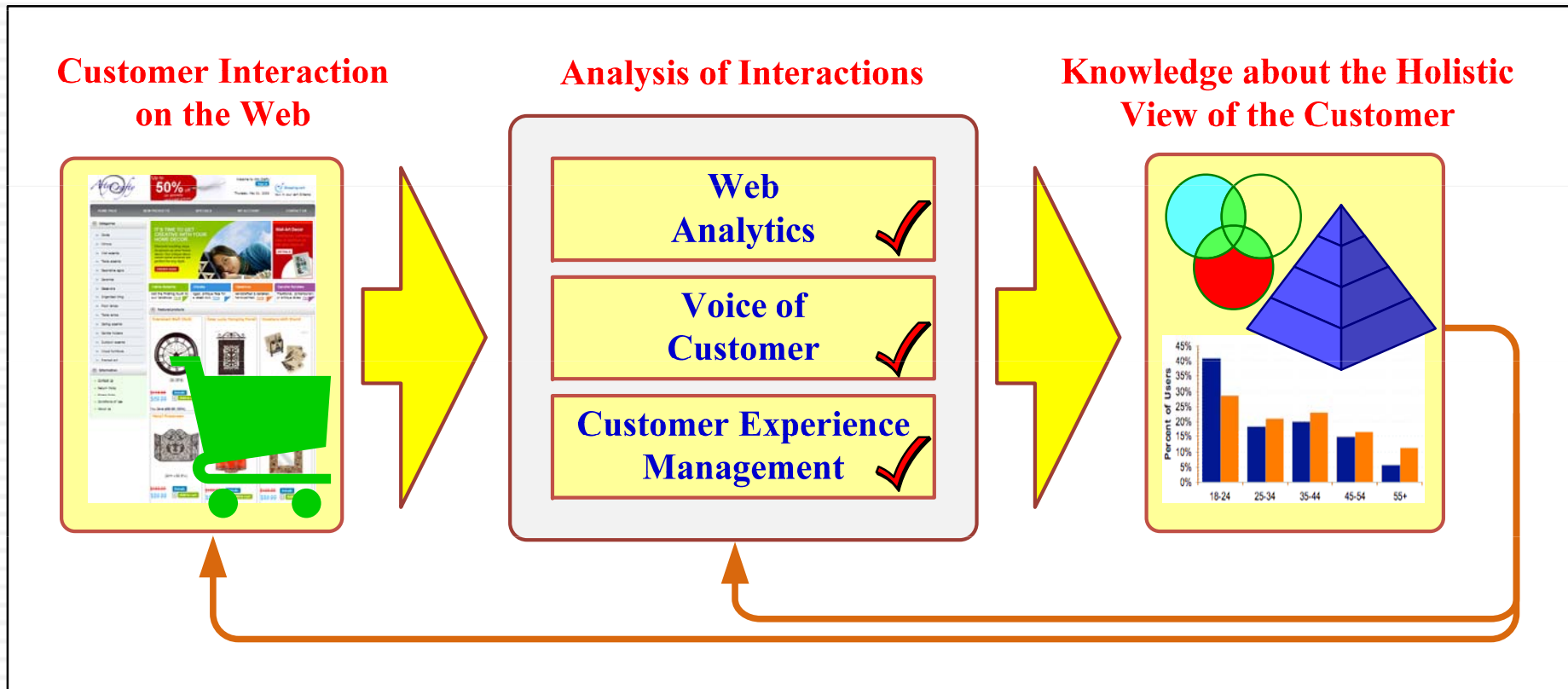
ارزیابی اقدامات ترفیعی

ارائه اطلاعات پویا به کاربران بر
اساس علایق و مشخصات آنها

پیش بینی رفتار کاربر بر اساس
قواعد و مشخصات شناسایی شده

اکوسیستم بهینه سازی وب سایت

nemati@ut.ac.ir



نرم افزارهای کاوش به کارگیری وب

nemati@ut.ac.ir

نام محصول	آدرس وبسایت	شرح محصول
Angoss Knowledge WebMiner	Angoss.com	تحلیل جریان کلیک
ClickTracks	Clicktracks.com	نمایش الگوهای بازدید کنندگان از وب سایت
LiveStats from DeepMetrix	Deepmetrix.com	تحلیل برخط بازدید، نمایش زنده در سایت
Megaputer WebAnalyst	http://megaputer.com/products/wm.php3	قابلیت های داده و متن کاوی
MicroStrategy Web Traffic Analysis Moudule	Microstrategy.com/solutions/Application/WTAM	گزارش حجم بازدیدها از سایت، تحلیل محتوا و گزارشات تحلیلی از بازدید وب
SAS Web Analytics	Sas.com/solution/webanalytics	تحلیل ترافیک وبسایت
SPSS Web Mining for Clementine	Spss.com/web_mining_for-clementine	استخراج رویدادهای وب

تمرین ۳-۹

45

nemati@ut.ac.ir

□ سه نرم افزار وب کاوی از سه شرکت مختلف را به دلخواه خود انتخاب کنید و قابلیت های آنها را با یکدیگر مقایسه نمایید.

تمرین ۹-۴

□ چند مقاله درباره وب کاوی مطالعه کرده و خلاصه ای از آنها را گزارش کنید.

- [1] E. Turban, R. Sharda, J. E. Aronson, and D. King (2011) “Business Intelligence A managing approach”, Second edition, Chapter 5.