کوییز اول

در استفاده از شاخص اصلی و کمکی در حالتی که مجموعه اسناد پویا (dynamic) باشد کدام موارد صحیح است؟

a. اگر لیست‌های پست‌ها در فایل‌های مجزایی ذخیره شوند فرآیند ادغام دو شاخص کارآمدتر است.

b. برای حذف یک سند از صفر کردن بیت آن سند در یک بردار بیتی استفاده می شود.

c. اگر کلیه لیست‌های پست‌ها در یک فایل واحد ذخیره شوند از نظر سیستم عامل بهتر است. ✔

d. در بدترین حالت یک posting به تعداد T بار در فرآیند ادغام شرکت می کند(T تعداد کل postingها است). ✖

The correct answers are:

برای حذف یک سند از صفر کردن بیت آن سند در یک بردار بیتی استفاده می شود.,

اگر لیست‌های پست‌ها در فایل‌های مجزایی ذخیره شوند فرآیند ادغام دو شاخص کارآمدتر است.,

اگر کلیه لیست‌های پست‌ها در یک فایل واحد ذخیره شوند از نظر سیستم عامل بهتر است.

کدام یک از جملات زیر در مورد Query refinement صحیح است؟

a. منظور اصلاح پرسمان (کوئری) ورودی بر اساس بازخورد دریافت شده از نتایج ارائه شده است. ✔

b. منظور اصلاح پرسمان توسط موتور جستجو به منظور اجرای سریع‌تر پرسمان کاربر در موتور جستجو است.

c. باعث کاهش تعداد نتایج بازگردانده شده می شود.

d. در مواردی که نتایج بازگردانده شده نیاز اطلاعاتی کاربر را برآورده نکند انجام می شود. ✔

The correct answers are:

منظور اصلاح پرسمان (کوئری) ورودی بر اساس بازخورد دریافت شده از نتایج ارائه شده است.,

در مواردی که نتایج بازگردانده شده نیاز اطلاعاتی کاربر را برآورده نکند انجام می شود.

کدام یک از موارد زیر از یک ماتریس وقوع کلمه-سند به شکل دقیق قابل استخراج است؟

- [x] a. لیست ایست واژه ها (stop words) ✔
- [ ] b. فرکانس یک کلمه در یک سند
- [ ] c. کلمات هم معنی
- [x] d. اسناد تکراری (اسناد دارای محتوای یکسان) ✖

پاسخ درست «
لیست ایست واژه ها (stop words)» است.

در معماری map-reduce برای ساخت شاخص برای یک مجموعه از اسناد

- [x] a. محل ذخیره سازی توکن های استخراج شده توسط هر پارسر مجزای از پارسرهای دیگر است. ✔
- [ ] b. هر پارسر توکن‌های یک زیر مجموعه مجزا از مجموعه اسناد ورودی را استخراج می کند.
- [ ] c. اینورترها وظیفه جمع آوری، مرتب سازی و ذخیره سازی زوج‌های (term,doc) در قالب لیست پست ها را برعهده دارد.
- [x] d. توکن های استخراج شده بر حسب حروف الفبا در زیربخش های جداگانه ای ذخیره می شوند. ✔

The correct answers are:
هر پارسر توکن‌های یک زیر مجموعه مجزا از مجموعه اسناد ورودی را استخراج می کند.,
توکن های استخراج شده بر حسب حروف الفبا در زیربخش های جداگانه ای ذخیره می شوند.,
محل ذخیره سازی توکن های استخراج شده توسط هر پارسر مجزای از پارسرهای دیگر است.,
اینورترها وظیفه جمع آوری، مرتب سازی و ذخیره سازی زوج‌های (term,doc) در قالب لیست پست ها را برعهده دارد.

کدام یک از موارد زیر یک متن نیمه-ساختارمند (semi-structured) محسوب می شود؟

- [ ] a. متن یک پیامک
- [ ] b. متن چکیده یک مقاله
- [ ] c. متن کوئری یک کاربر در حالت free text
- [x] d. متن موجود در یک صفحه وب ✔

پاسخ درست «
متن موجود در یک صفحه وب» است.

در الگوریتم BSBI: Blocked sort-based Indexing

- [x] a. برای افزایش سرعت می توان داده های ورودی را از دو کپی از داده ها در دو دیسک مجزا خواند و به طور همزمان در دو دیسک مجزای دیگر نوشت.
- [ ] b. فقط برای ایجاد شاخص غیرمکانی قابل استفاده است.
- [ ] c. با هر بار تکرار بخش ادغام الگوریتم به شرط زوج بودن تعداد بخش های مرتب، تعداد این بخش های مرتب شده نصف و اندازه هر یک از آنها دو برابر می شود.
- [ ] d. برای هر بخش مرتب شده از اسناد دیکشنری مجزایی ایجاد می شود.

The correct answers are:

با هر بار تکرار بخش ادغام الگوریتم به شرط زوج بودن تعداد بخش های مرتب، تعداد این بخش های مرتب شده نصف و اندازه هر یک از آنها دو برابر می شود.,

برای افزایش سرعت می توان داده های ورودی را از دو کپی از داده ها در دو دیسک مجزا خواند و به طور همزمان در دو دیسک مجزای دیگر نوشت.

کدام یک از موارد زیر در مورد شاخص مکانی (positional index) صحیح است؟

- [ ] a. ممکن است در تشخیص پرسمان های عبارتی (phrase query) با استفاده از این شاخص خطای مثبت اشتباه (false positive) داشته باشیم.
- [x] b. با استفاده از آن می توان به جستجوهای مجاورتی (proximity search) جواب داد.
- [x] c. به ازاء هر کلمه کلیه محل‌های وقوع آن کلمه در کلیه اسناد ذخیره می شود.
- [ ] d. اندازه آن 2 تا 4 برابر حجم اسناد اولیه است.

The correct answers are:

به ازاء هر کلمه کلیه محل‌های وقوع آن کلمه در کلیه اسناد ذخیره می شود.,

با استفاده از آن می توان به جستجوهای مجاورتی (proximity search) جواب داد.

لیست‌های پست های (postings lists) دو کلمه دارای طول‌های (merge) . پیچیدگی عملیات ادغام x<y هستند که y و x برای یافتن اسناد مشترک در لیست پست‌های این دو کلمه در بهترین حالت کدام است؟

- a. x ✔
- b. y
- c. x+y
- d. y-x

پاسخ درست « x» است.

در ذخیره و بازیابی اطلاعات بر روی دیسک سخت (hard disk) :

- a. خواندن و نوشتن اطلاعات در قالب کلمه (ترکیبی از چند بایت) انجام می شود.
- b. بیشترین بخش از زمان خواندن یک بلاک زمان انتقال داده(transfer time) است.
- c. خواندن و نوشتن اطلاعات در قالب بلاک انجام می شود.. ✔
- d. بیشترین بخش از زمان خواندن یک بلاک زمان استوانه یابی (seek time) است. ✔

The correct answers are:

بیشترین بخش از زمان خواندن یک بلاک زمان استوانه یابی (seek time) است.,

خواندن و نوشتن اطلاعات در قالب بلاک انجام می شود..

کوییز دوم

کدام یک از موارد زیر از اهداف فشرده سازی در موتورهای بازیابی اطلاعات است؟

- [x] a. افزایش سرعت انتقال اطلاعات از حافظه جانبی (دیسک) به حافظه اصلی ✔
- [ ] b. افزایش سرعت انتقال اطلاعات نتایج از موتور جستجو به سیستم کاربر
- [x] c. کاهش حجم حافظه مصرفی ✔
- [ ] d. افزایش سرعت انتقال اطلاعات از حافظه اصلی به پردازنده (CPU)

The correct answers are:
افزایش سرعت انتقال اطلاعات از حافظه جانبی (دیسک) به حافظه اصلی,
کاهش حجم حافظه مصرفی

طبق قانون هیپس (heaps) کدام یک از موارد زیر صحیح است؟

- [ ] a. نمودار log-log اندازه دیکشنری در مقابل تعداد توکن ها نیم ساز زاویه بین محورهای مختصات می شود.
- [x] b. با پردازش بخشی از اسناد ورودی، از این قانون می توان برای پیش بینی اندازه دیکشنری به ازاء تعداد زیاد توکن استفاده کرد. ✔
- [x] c. هر چه تعداد توکن های مجموعه اسناد بیشتر باشد اندازه دیکشنری آنها نیز بزرگتر است. ✔
- [ ] d. اندازه دیکشنری دارای یک رابطه خطی با تعداد توکن ها است.

The correct answers are:
هر چه تعداد توکن های مجموعه اسناد بیشتر باشد اندازه دیکشنری آنها نیز بزرگتر است.,
با پردازش بخشی از اسناد ورودی، از این قانون می توان برای پیش بینی اندازه دیکشنری به ازاء تعداد زیاد توکن استفاده کرد.

در ذخیره سازی دیکشنری به صورت رشته (dictionary as a string) طول اشاره گرهای ترم بر چه اساسی مشخص می شود؟

- ○ a. تعداد کلمات دیکشنری
- ◉ b. طول رشته حاصل از کل کلمات دیکشنری ✔
- ○ c. تعداد توکن های اسناد پردازش شده
- ○ d. حجم اسناد پردازش شده

پاسخ درست «
طول رشته حاصل از کل کلمات دیکشنری» است.

یک دیکشنری متشکل از پنج کلمه W1 W2 W3 W4 W5 را در نظر بگیرید. برای یافتن یک کلمه در دیکشنری از جستجوی باینری استفاده می شود. در صورتی که نسبت تعداد دفعات مراجعه برای یافتن هر یک از کلمات به ترتیب 0.1,0.1,0.2,0.3,0.3 باشد، متوسط تعداد مقایسه ها برای یافتن کلمات در دیکشنری چند است؟

جواب را به صورت یک عدد با حداکثر یک رقم اعشار و به صورت رقمی و به انگلیسی در کادر مربوط به جواب وارد نموده و از درج هر گونه متن اضافی یا فاصله خودداری کنید.

Answer: 0.5 ✖

پاسخ درست: 2.2

کد گاما مربوط به دنباله اسناد 5,6,8 چیست؟

شماره های مشخص شده شماره اسنادی هستندکه یک کلمه در آنها آمده است و postings list یک کلمه را تشکیل می دهند.

از روش کدگذاری فاصله (gap) استفاده می شود.

پاسخ خود را به صورت یک رشته باینری صرفا با استفاده از ارقام 0 و 1 که به صورت انگلیسی نوشته می شوند در کادر پاسخ وارد کرده و از اضافه کردن هر گونه کاراکتر اضافه خودداری کنید.

Answer: 100010100 ✖

پاسخ درست: 110010100

در روش شاخص جایگشتی (permuterm index)، به ازاء یک کلمه شش حرفی چند کلمه در درخت B مربوطه درج می شود؟

- a. 7 ✔
- b. 6
- c. 1
- d. به تعداد جایگشت های ممکن شش حرف کلمه با حذف حالات تکراری

پاسخ درست « 7» است.

کدام یک از جملات زیر در مورد انواع خطاهای املایی و روش های اصلاح خطا درست است؟

ترجمه برخی کلمات مرتبط جهت درک بهتر گزینه ها:

غیر کلمه ای : non-word

کلمه واقعی : real-word

حساس به زمینه: context sensitive

غیرحساس به زمینه: context insensitive

- a. معمولا تشخیص خطاهای غیر کلمه ای نیازمند استفاده از زمینه است.
- b. معمولا خطاهای غیرکلمه ای غیرحساس به زمینه هستند. ✔
- c. اصلاح خطاهای کلمه واقعی عمدتا نیاز به استفاده از زمینه دارد.
- d. معمولا تشخیص خطاهای کلمه واقعی نیازمند استفاده از زمینه است. ✔

The correct answers are:

معمولا خطاهای غیرکلمه ای غیرحساس به زمینه هستند., اصلاح خطاهای کلمه واقعی عمدتا نیاز به استفاده از زمینه دارد.,

معمولا تشخیص خطاهای کلمه واقعی نیازمند استفاده از زمینه است.

در مدل کانال نویزی برای اصلاح خطا، از قاعده ی بیز برای ..... استفاده می شود.

- [ ] a. محاسبه احتمال ارسال کلمه w در ورودی کانال به شرط مشاهده کلمه x در خروجی کانال
- [x] b. دیکد کردن کلمه خروجی از کانال ✔
- [x] c. کد کردن کلمه ورودی به کانال ✖
- [x] d. محاسبه احتمال مشاهده کلمه x در خروجی کانال به شرط ارسال کلمه w در ورودی کانال ✖

The correct answers are:
دیکد کردن کلمه خروجی از کانال,
محاسبه احتمال ارسال کلمه w در ورودی کانال به شرط مشاهده کلمه x در خروجی کانال

منظور از فاصله ویرایشی لونشتاین دو کلمه چیست؟

- a. حداکثر احتمال تبدیل یک کلمه به کلمه دیگر
- b. حداقل احتمال تبدیل یک کلمه به کلمه دیگر
- c. حداقل تعداد ویرایش های مورد نیاز برای تبدیل یک کلمه به کلمه دیگر ✔
- d. حداکثر تعداد ویرایش‌های مورد نیاز برای تبدیل یک کلمه به کلمه دیگر

پاسخ درست «

حداقل تعداد ویرایش های مورد نیاز برای تبدیل یک کلمه به کلمه دیگر» است.

کدام یک از موارد زیر در مورد مدل زبانی بایگرام درست است؟

- [x] a. می توان از روش افزودن یک (add-1) برای هموارسازی آن استفاده کرد. ✔
- [x] b. احتمال مورد نظر برای مدل کانال نویزی در رابطه بیز را فراهم می کند. ✖
- [x] c. می توان از ترکیب آن با مدل زبانی یونیگرام برای هموارسازی آن استفاده کرد. ✔
- [x] d. احتمال وقوع یک کلمه را مشروط به کلمه قبل از آن فراهم می کند. ✔

The correct answers are:
احتمال وقوع یک کلمه را مشروط به کلمه قبل از آن فراهم می کند.,
می توان از روش افزودن یک (add-1) برای هموارسازی آن استفاده کرد.,
می توان از ترکیب آن با مدل زبانی یونیگرام برای هموارسازی آن استفاده کرد.

کوییز سوم

شباهت دو سند A و B بر اساس ضریب جکارد 0.4 و شباهت دو سند C و D بر اساس ضریب جکارد 0.7 است. در صورتی که متن سند C را به انتهای سند A و متن سند D را به انتهای سند B اضافه کنیم، میزان شباهت دو سند حاصل بر اساس معیار جکارد کدام یک از موارد زیر می تواند باشد؟

- [x] a. بین 0.4 تا 0.7 ✔
- [ ] b. صفر
- [ ] c. بیشتر از 0.7
- [ ] d. کمتر از 0.4

The correct answers are:

کمتر از 0.4,

بین 0.4 تا 0.7,

بیشتر از 0.7

اگر کلمات کوئری q همگی در سند A وجود داشته باشند، شباهت کسینوسی بردارهای کوئری q و سند A در مدل lnc.ltc کدام است؟

- [ ] a. هر عددی در بازه صفر تا یک می تواند باشد.
- [ ] b. دقیقا برابر با یک خواهد بود.
- [x] c. قطعا عددی غیر صفر (بزرگتر از صفر) خواهد بود. ✖
- [ ] d. در صورتی که سند A هیچ کلمه دیگری بجز کلمات کوئری نداشته باشد شباهت کسینوسی سند و کوئری برابر با یک خواهد شد.

پاسخ درست «

هر عددی در بازه صفر تا یک می تواند باشد.» است.

بهره تجمعی کاهشی نرمال شده (Normalized Discounted Cumulative Gain) برای یک موتور بازیابی اطلاعات را در حالتی که سه سند مرتبط با میزان ارتباط 3، 2 و 1 وجود داشته باشد و ترتیب نتایج موتور بازیابی اطلاعات مورد نظر به گونه ای باشد که سند اول دارای میزان ارتباط 2، سند دوم دارای میزان ارتباط 1 و سند سوم دارای میزان ارتباط 3 باشد چقدر است؟

Log2=1, log3=1.6

Answer: 0.86 ✔

پاسخ درست: 0.866

کدام یک از گزینه های زیر در مورد روش وزن دهی کلمات tf-idf درست است؟

a. این که یک کلمه در چه تعداد سند آمده باشد نسبت به این که این کلمه در کل مجموعه اسناد چند دفعه تکرار شده است مهمتر است.

b. وزن هر کلمه در هر سند با تعداد تکرار آن کلمه در آن ستند مرتبط است ✔

c. وزن هر کلمه در هر سند (میزان ارتباط یک کلمه با یک سند) الزاما با تعداد تکرار آن کلمه در آن سند متناسب نیست.

d. هر چه یک کلمه در اسناد کمتری آمده باشد، آن کلمه دارای اطلاعات مفید بیشتری است. ✔

The correct answers are:

وزن هر کلمه در هر سند با تعداد تکرار آن کلمه در آن ستند مرتبط است,

وزن هر کلمه در هر سند (میزان ارتباط یک کلمه با یک سند) الزاما با تعداد تکرار آن کلمه در آن سند متناسب نیست.,

هر چه یک کلمه در اسناد کمتری آمده باشد، آن کلمه دارای اطلاعات مفید بیشتری است.,

این که یک کلمه در چه تعداد سند آمده باشد نسبت به این که این کلمه در کل مجموعه اسناد چند دفعه تکرار شده است مهمتر است.

مهمترین معیار در ارزیابی نتایج یک موتور جستجو چیست؟

a. میزان کلیک نتایج توسط کاربر

b. شباهت کسینوسی با کوئری کاربر

c. در بر داشتن کلمات کوئری

d. برطرف کردن نیاز اطلاعاتی کاربر ✔

پاسخ درست «

برطرف کردن نیاز اطلاعاتی کاربر» است.

کدام یک از موارد زیر از ویژگی های ارزیابی های انسانی برای تشخیص ارتباط کوئری ها با اسناد است؟

a. ناسازگاری در میان ارزیابی های افراد مختلف ✔

b. هزینه زیاد ✔

c. ناسازگاری ارزیابی های یک فرد در طول زمان ✔

در طول زمان ناسازگار است

d. عدم درک مفهوم کوئری توسط کاربر

The correct answers are:

هزینه زیاد,

ناسازگاری در میان ارزیابی های افراد مختلف
,

ناسازگاری ارزیابی های یک فرد در طول زمان

d. عدم درک مفهوم کوئری توسط کاربر

The correct answers are:

هزینه زیاد,

ناسازگاری در میان ارزیابی های افراد مختلف
,

ناسازگاری ارزیابی های یک فرد در طول زمان

در طول زمان ناسازگار است,
عدم درک مفهوم کوئری توسط کاربر

بردارهای با طول واحد (نرمال شده) d1 و d2 بازنمائی برداری دو سند با استفاده از روش tf-idf هستند. در صورتی که این دو سند هیچ کلمه مشترکی نداشته باشند، کدام یک از گزینه های زیر درست است؟

a. فاصله اقلیدسی دو سند یک است.

b. فاصله اقلیدسی دو سند 1.4 (جذر دو) است. ✔

c. شباهت کسینوسی دو سند یک است.

d. شباهت کسینوسی دو سند صفر است. ✔

The correct answers are:

شباهت کسینوسی دو سند صفر است.,

فاصله اقلیدسی دو سند 1.4 (جذر دو) است.

پاسخ های یک موتور جستجو برای دو کوئری q1 و q2 به شکل زیر است. ترتیب نتایج از چپ به راست است. نتایج مرتبط با علامت R و نتایج غیرمرتبط با علامت N نشان داده شده است. در این حالت مقدار میانگین رتبه متقابل Mean Reciprocal Rank را محاسبه کنید.

q1:NRNRR

q2:NNNRR

Answer: 0.375 ✔

پاسخ درست: 0.375

کوییز چهارم

برای بهبود سرعت محاسبه امتیاز اسناد از خوشه بندی استفاده می کنیم. در صورتی که تعداد کل اسناد N=100 سند و تعداد خوشه ها k=20 خوشه باشد و هر دنبال کننده (سند) b1=2 مرکز خوشه را دنبال کند و در زمان جستجو هر سند با b2=4 خوشه مقایسه شود. برای پاسخگویی به یک کوئری حداکثر چند مقایسه برداری (اندازه گیری شباهت یا فاصله دو بردار) انجام می شود؟

- a. 80 ✔
- b. 60
- c. 40
- d. 20

The correct answers are:
20,
40,
60,
80

کدام یک از موارد زیر جزء تکنیک های حذف شاخص (index elimination) برای افزایش سرعت امتیازدهی اسناد نیست؟

- a. محاسبه شباهت برای کلماتی که دارای وزن بیشتری در سند هستند ✔
- b. انتخاب اسنادی که حداقل یک کلمه از کلمات کوئری را داشته باشند.
- c. انتخاب اسنادی که حداقل چند کلمه از کلمات کوئری را داشته باشند.
- d. محاسبه شباهت برای کلمات با idf بیشتر.

پاسخ درست «
محاسبه شباهت برای کلماتی که دارای وزن بیشتری در سند هستند» است.

در روش دسته بندی k نزدیک ترین همسایه (KNN) هر چه اندازه k بیشتر شود.

- [x] a. مرز بین کلاس ها هموارتر می شود. ✔
- [x] b. بایاس بیشتر می شود. ✔
- [ ] c. حساسیت به نویز کمتر می شود.
- [x] d. واریانس مدل کاهش می یابد. ✔

The correct answers are:

مرز بین کلاس ها هموارتر می شود.,

بایاس بیشتر می شود. ,

حساسیت به نویز کمتر می شود.,

واریانس مدل کاهش می یابد.

کدام یک از موارد زیر در مورد انتخاب ویژگی (feature selection) درست است؟

- [x] a. بهبود قابلیت تعمیم (generalization) مدل
- [x] b. کوچکتر کردن دسته بند از نظر حافظه مورد نیاز
- [ ] c. عدم امکان استفاده از برخی دسته بندها در صوت عدم انتخاب ویژگی
- [x] d. افزایش سرعت اجرای الگوریتم

The correct answers are:
افزایش سرعت اجرای الگوریتم, عدم امکان استفاده از برخی دسته بندها در صوت عدم انتخاب ویژگی,
کوچکتر کردن دسته بند از نظر حافظه مورد نیاز,
بهبود قابلیت تعمیم (generalization) مدل

در یک شبکه اجتماعی که افراد اقدام به ارسال نظرات خود در مورد مسائل مختلف می کنند، و می توانند نظرات دیگران را لایک کنند، کدام یک از گزینه های زیر معیار مناسبی برای استفاده به عنوان یک امتیاز استاتیک برای نظرات ارسالی کاربران نیست؟ نامناسب ترین معیار را انتخاب کنید.

- a. تعداد دنبال کننده ها (فالورها) کاربر ارسال کننده نظر
- b. تعداد لایک های قبلی کاربر ارسال کننده نظر
- c. تعداد نظرات قبلی ارسال شده توسط کاربر
- [x] d. تعداد لایک های پست

پاسخ درست «
تعداد نظرات قبلی ارسال شده توسط کاربر» است.

کدام یک از توابع زیر یک تابع مناسب برای محاسبه امتیاز کل (net score) مرکب از امتیاز شباهت کسینوسی (c) و امتیاز اعتبار (g) نیست؟

- a. g+c
- b. g به توان c
- c. g*c
- [x] d. g/c

The correct answers are:
g/c,
g به توان c

برای دسته بندی یک سند به C کلاس ممکن از یک دسته بند بیز ساده با k ویژگی باینری (دو حالته) استفاده می کنیم. این مدل دارای چند پارامتر (مقدار احتمال) است که می بایست از روی داده های آموزشی محاسبه شوند؟

لازم نیست فقط پارامترهای آزاد را در نظر بگیرید. کلیه پارامترها در نظر گرفته شوند.

a. C(2k+1)

b. 2K+C

c. C(2k-1)

d. k+C ✖

پاسخ درست «C(2k+1)» است.

کدام یک از موارد زیر در مورد الگوریتم WAND صحیح است؟

جستجو با سه کلمه t2 ،t1 و t3 انجام می شود.

مقدار باند بالا برای کلمات به ترتیب ub2 ،ub1 و ub3 است.

نشانگر کلمات با ptr2 ،ptr1 و ptr3 نمایش داده می شود.

شماره مربوط به اسناد در محل نشانگر کلمات doc_id2 ،doc_id1 و doc_id3 هستند. همچنین doc_id1<doc_id2 <docid_3 است.

حد آستانه (threshold) فعلی الگوریتم برابر با T است.

a. به ازاء هر کلمه مقادیر باند بالا کاهشی است. یعنی ub(ptr) >= ub(ptr+1)

b. با توجه به این که doc_id_1<doc_id2 پس ub1<ub2

c. اگر T > ub1+ub2+ub3 باشد الگوریتم خاتمه می یابد و بقیه اسناد بررسی نمی شوند. ✔

d. اگر T< ub1 باشد هیچ سندی در این مرحله هرس نمی شود. ✔

The correct answers are:
به ازاء هر کلمه مقادیر باند بالا کاهشی است. یعنی ub(ptr) >= ub(ptr+1),
اگر T > ub1+ub2+ub3 باشد الگوریتم خاتمه می یابد و بقیه اسناد بررسی نمی شوند.,
اگر T< ub1 باشد هیچ سندی در این مرحله هرس نمی شود.

کدام موارد در رابطه با روش دسته بندی مبتنی بر قانون (rule based) درست است؟

a. در سازمان های دولتی و شرکت های بزرگ کمتر مورد استفاده قرار می گیرد.

b. هزینه بالایی دارد. ✔

c. نسبت به روش دستی نیاز به برچسب زنی داده های کمتری دارد.

d. دقت آن نسبت به حالت برچسب زنی دستی ممکن است کمتر یا بیشتر باشد. ✔

The correct answers are:
نسبت به روش دستی نیاز به برچسب زنی داده های کمتری دارد.,
هزینه بالایی دارد.,
دقت آن نسبت به حالت برچسب زنی دستی ممکن است کمتر یا بیشتر باشد.

کدام یک از موارد زیر در مورد کوئری های مانا (standing query) صحیح است؟

- [x] a. به دفعات زیاد در موتور بازیابی اطلاعات جستجو می شوند. ✔
- [ ] b. کوئری در جستجوهای مختلف تقریبا ثابت باقی می ماند.
- [x] c. کاربر در هر بار جستجو انتظار دریافت نتایج جدیدی دارد. ✔
- [x] d. بیشتر بیانگر یک مساله دسته بندی است تا رتبه بندی. ✔

The correct answers are: به دفعات زیاد در موتور بازیابی اطلاعات جستجو می شوند.,
کوئری در جستجوهای مختلف تقریبا ثابت باقی می ماند.,
کاربر در هر بار جستجو انتظار دریافت نتایج جدیدی دارد.,
بیشتر بیانگر یک مساله دسته بندی است تا رتبه بندی.