کد کنترل

334

F

334F

آزمون (نیمه‌متمرکز) ورود به دوره‌های دکتری ـ سال ۱۴۰۱

«اگر دانشگاه اصلاح شود مملکت اصلاح می‌شود.»
امام خمینی (ره)

دفترچهٔ شمارهٔ (۱)
صبح جمعه ۱۴۰۰/۱۲/۶

جمهوری اسلامی ایران
وزارت علوم، تحقیقات و فنّاوری
سازمان سنجش آموزش کشور

رشتهٔ مهندسی فنّاوری اطلاعات
(کد ۲۳۵۸)

جدول مواد امتحانی، تعداد، شمارهٔ سؤال‌ها و زمان پاسخ‌گویی

مواد امتحانی	تعداد سؤال	از شمارهٔ	تا شمارهٔ	زمان پاسخ‌گویی
مجموعه دروس تخصصی: ـ حل مسئله (ساختمان داده‌ها و طراحی الگوریتم‌ها) ـ شبکه (شبکه‌های کامپیوتری و امنیت شبکه) ـ مهندسی اطلاعات (پایگاه داده پیشرفته، بازیابی پیشرفته اطلاعات و داده‌کاوی) ـ خدمات فنّاوری اطلاعات (تجارت الکترونیک و آموزش الکترونیکی)	۴۵	۱	۴۵	۱۵۰ دقیقه

این آزمون نمرهٔ منفی دارد.

استفاده از ماشین‌حساب مجاز نیست.

* متقاضی گرامی، وارد نکردن مشخصات و امضا در کادر زیر، به منزلهٔ غیبت و حضور نداشتن در جلسهٔ آزمون است.

اینجانب ..... با شمارهٔ داوطلبی ..... با آگاهی کامل، یکسان‌بودن شمارهٔ صندلی خود را با شمارهٔ داوطلبی مندرج در بالای کارت ورود به جلسه، بالای پاسخ‌نامه و دفترچهٔ سؤال‌ها، نوع و کد کنترل درج شده بر روی دفترچهٔ سؤال‌ها و پایین پاسخ‌نامه‌ام را تأیید می‌نمایم.

امضا:

۱- اگر بخواهیم داده‌ساختار صف را با استفاده از پشته پیاده‌سازی کنیم، طوری‌که عملیات‌های پایه‌ای صف در زمان سرشکن $O(1)$ قابل انجام باشد، کدام مورد درست است؟

۱) با سه پشته می‌توان این کار را انجام داد و سه پشته برای این کار لازم است.

۲) با دو پشته می‌توان این کار را انجام داد و دو پشته برای این کار لازم است.

۳) با یک پشته می‌توان این کار را انجام داد.

۴) این کار امکان‌پذیر نیست.

۲- فرض کنید m آرایهٔ مرتب داریم که در مجموع آرایه‌ها شامل n عدد هستند. می‌خواهیم از هر کدام از آرایه‌ها یک عدد را انتخاب کنیم، به‌طوری‌که اختلاف بیشینه و کمینهٔ اعداد انتخاب‌شده کمترین مقدار ممکن شود. این کار در چه زمانی ممکن است؟ (بهترین گزینه را انتخاب کنید.)

۱) $O(m+\log n)$ ۲) $O(n\log m)$

۳) $O(n+m)$ ۴) $O(nm)$

۳- نمایش پیش‌ترتیب (Preorder) و میان‌ترتیب (Inorder) یک درخت دودویی مرتب با n رأس داده شده است. کدام مورد درست است؟

۱) درخت متناظر با این لیست رئوس، منحصربه‌فرد است.

۲) دقیقاً دو درخت با این ترتیب ملاقات رئوس وجود دارد.

۳) دقیقاً $\frac{n}{2}$ درخت با این ترتیب ملاقات رئوس وجود دارد.

۴) هیچ یک از موارد درست نیست.

۴- فرض کنید یک هرم بیشینه شامل حداکثر n عدد داده شده است. جستجوی یک مقدار در این هرم بیشینه در کدام کلاس پیچیدگی است؟ (بهترین گزینه را انتخاب کنید.)

۱) $O(\frac{n}{\log(n)})$ ۲) $O(\log(n))$ ۳) $O(n)$ ۴) $O(\sqrt{n})$

۵- کدام مورد درخصوص دو گزارهٔ (الف) و (ب) به‌ترتیب درست است؟

الف ـ اگر $f(n)\in O(2^{2n})$، آنگاه $f(n)\in \Omega(2^n)$.

ب ـ اگر $T(n)=2T(\frac{n}{2})+O(n^2)$، آنگاه $T(n)=\Theta(n^2)$.

۱) درست ـ درست ۲) درست ـ نادرست

۳) نادرست ـ درست ۴) نادرست ـ نادرست

۶- رشته‌ای به طول اولیهٔ $m \geq 1$ در یک صفحه از نرم‌افزار حروف‌چین داریم. دو عمل زیر را می‌توانیم به‌ترتیب دلخواه روی این رشته انجام دهیم.

- کپی: کل رشتهٔ موجود در صفحه را در حافظه ذخیره کن.
- پیست: رشتهٔ ذخیره شده در حافظه را به انتهای رشتهٔ موجود در صفحه اضافه کن.

به‌طور مثال، اگر رشتهٔ اولیه **ab** باشد، پس از انجام یک عمل کپی و یک عمل پیست، رشتهٔ موجود در صفحه به‌صورت **abab** درخواهد آمد. اگر **dp[n]** نشان‌دهندهٔ طول بزرگ‌ترین رشتهٔ قابل ایجاد با **n** بار استفاده از عمل کپی یا پیست (به‌ترتیب دلخواه) باشد، کدام رابطهٔ بازگشتی زیر برای محاسبهٔ **dp[n]** به‌ازای $n \geq 5$ درست است؟ (فرض کنید در ابتدا حافظه خالی است.)

۱) $dp[n] = 2dp[n-2]$
۲) $dp[n] = 2dp[n-1]$
۳) $dp[n] = \max(2dp[n-2], 3dp[n-3])$
۴) $dp[n] = \max(dp[n-1], 2dp[n-2])$

۷- محاسبهٔ عنصر با بیشترین تکرار در یک آرایهٔ **n** عضوی دلخواه در چه زمانی ممکن است؟ (بهترین گزینه را انتخاب کنید.)

۱) $O(n\log(n))$
۲) $O(n\sqrt{n})$
۳) $O(n^2)$
۴) $O(n)$

۸- کدام‌یک از دو گزارهٔ (الف) و (ب) درخصوص الگوریتم هافمن به‌ترتیب درست است؟

الف ـ اگر فراوانی یک حرف در یک متن بیشتر از ۰/۴ باشد، آنگاه درخت هافمن متن لزوماً شامل کد یک بیتی است.

ب ـ اگر فراوانی هر حرف در یک متن کمتر از ۰/۳ باشد، آنگاه درخت هافمن متن لزوماً شامل کد یک بیتی نیست.

۱) نادرست ـ درست ۲) درست ـ درست ۳) درست ـ نادرست ۴) نادرست ـ نادرست

۹- فرض کنید **G** یک گراف کامل وزن‌دار با **n** رأس است. درخت پوشای کمینهٔ **G** در چه زمانی قابل محاسبه است؟ (بهترین گزینه را انتخاب کنید.)

۱) $\Theta(n)$
۲) $\Theta(n\log n)$
۳) $\Theta(n^2\log n)$
۴) $\Theta(n^2)$

۱۰- آرایهٔ مرتب **A** شامل **n** عدد به‌صورت اکیداً صعودی داده شده است. یک نفر این آرایه را به‌اندازه **k** واحد شیفت دوری داده و نتیجه را به‌صورت یک آرایهٔ **B** به ما داده است. هدف پیدا کردن مقدار **k** است. در چه زمانی می‌توان مقدار **k** را با داشتن آرایهٔ **B** محاسبه کرد؟ (بهترین گزینه را انتخاب کنید.)

۱) $O(n)$
۲) $O(\sqrt{n})$
۳) $O(\log n)$
۴) $O(\log^2 n)$

۱۱- به‌ازای اعداد صحیح مثبت **n** و **m** و **a**، در چه مرتبهٔ زمانی می‌توان $a^n \bmod m$ را محاسبه کرد؟ فرض کنید عملیات‌های ضرب و جمع به پیمانهٔ **m** در زمان $O(1)$ قابل انجام هستند. (بهترین گزینه را انتخاب کنید.)

۱) $O(\log n)$
۲) $O(\sqrt{n})$
۳) $O(n)$
۴) $O(m)$

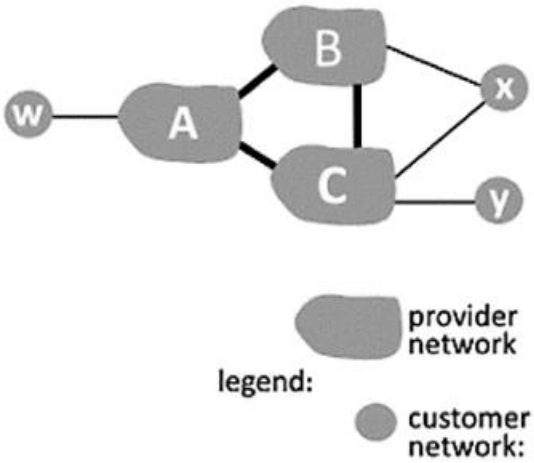
۱۲- کدام گزینه درخصوص دو گزاره (الف) و (ب) به‌ترتیب درست است؟

الف ـ به‌ازای هر گراف **G**، ترتیبی از یال‌های گراف وجود دارد که با تنها یک دور ریلکس کردن یال‌ها به آن ترتیب در اجرای الگوریتم بلمن ـ فورد، کوتاه‌ترین مسیر از رأس **s** به تمام رأس‌های دیگر محاسبه می‌شوند.

ب ـ در مرتب‌سازی ادغامی هر عنصر با $O(\log n)$ عنصر دیگر مقایسه می‌شود.

۱) درست ـ درست
۲) درست ـ نادرست
۳) نادرست ـ درست
۴) نادرست ـ نادرست

۱۳- فرض کنید یک ارائه‌دهندهٔ شبکه بخواهد ترافیک مربوط به مشتریان خود را منتقل کند. با توجه به شکل زیر، شبکه C مسیر Cy را به کدام موجودیت‌ها آگهی می‌کند؟



۱) فقط به x

۲) فقط به x و A

۳) فقط به A و B و x

۴) به تمام موجودیت‌ها

۱۴- دو کلمه ۱۶ بیتی 11010011 11110101 و 01000100 10110011 را در نظر بگیرید. کدام گزینه برای وارسی مجموع اینترنت این دو کلمه درست است؟

۱) 10101001 00010111

۲) 01010110 11100111

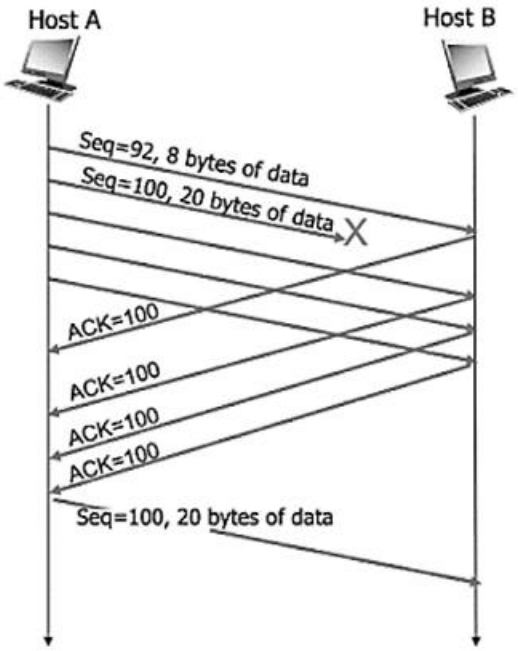
۳) 01010110 11101000

۴) 01011110 11000101

۱۵- بهینه‌سازی باز ارسال سریع TCP را که در شکل زیر به تصویر کشیده شده در نظر بگیرید. در نظر داشته باشید که فرستنده مطمئن نیست که سگمنت با شماره ترتیب 100 در واقع تلف شده است. آیا یک فرستنده می‌تواند ۳ ACK تکراری برای یک سگمنتی که در واقع تلف نشده است، دریافت کند؟ با توجه به این مطلب کدام یک از عبارات درست است؟ (در مدل کانال در نظر گرفته‌شده بسته‌ها می‌توانند تلف شوند ولی دچار خرابی نخواهند شد.)

الف) اگر کانال بتواند ترتیب بسته‌ها را عوض کند، امکان ACK تکراری سه‌گانه حتی اگر بسته‌ای تلف نشده باشد، وجود دارد.

ب) اگر کانال نتواند ترتیب بسته‌ها را عوض کند، ACK تکراری سه‌گانه در فرستنده نشانه آن است که حتماً یک بسته تلف شده است.



۱) هر دو عبارت

۲) عبارت الف

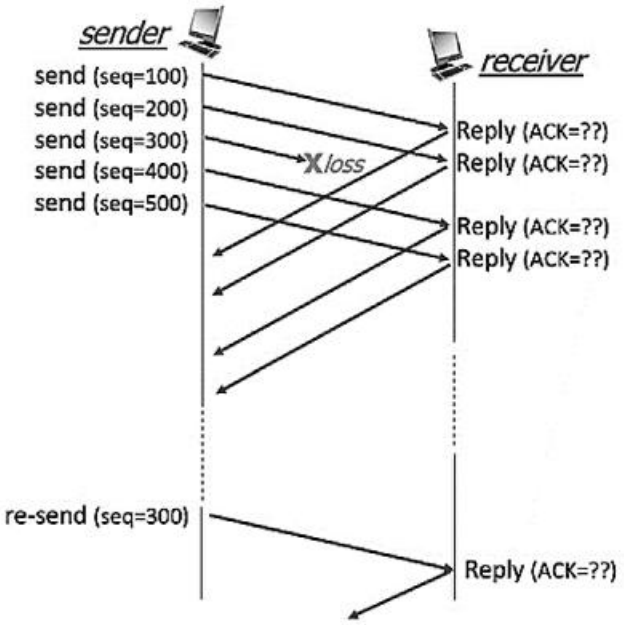
۳) عبارت ب

۴) علاوه بر اطلاعات فوق، فرستنده نیازمند اطلاعات اضافی است.

۱۶- **با توجه به شکل زیر، فرستنده TCP سگمنت‌هایی که قسمت محموله آن‌ها 100 بایتی است را ارسال می‌کند. فرستنده TCP پنج قطعه با شماره‌های به‌ترتیب 100، 200، 300، 400 و 500 ارسال می‌کند. فرض کنید سگمنت با شماره ترتیب 300 تلف شده باشد. گیرنده TCP بسته‌های دریافتی را بافر می‌کند. عبارات زیر را در نظر بگیرید:**

* **پس از دریافت سگمنت 100 گیرنده ACK با مقدار A را ارسال می‌کند.**
* **پس از دریافت سگمنت 200 گیرنده ACK با مقدار B را ارسال می‌کند.**
* **پس از دریافت سگمنت 500 گیرنده ACK با مقدار C را ارسال می‌کند.**
* **پس از دریافت بسته باز ارسالی 300 گیرنده ACK با مقدار D را ارسال می‌کند.**
* **گیرنده TCP پاسخی با ACK به مقدار E را ارسال نمی‌کند.**

با توجه به موارد فوق کدام گزینه درست است؟



۱) A=100, B=200, C=400, D=500, E=600

۲) A=200, B=300, C=300, D=400, E=600

۳) A=100, B=200, C=200, D=500, E=300

۴) A=200, B=300, C=300, D=600, E=400

۱۷- **کدام‌یک از موارد جزو مؤلفه‌های اساسی پروتکل IPv4 است؟**

۱) قالب دیتاگرام IPv4 ـ پروتکل‌ها و الگوریتم‌های مسیریابی (مثل OSPF و BGP) ـ قواعد مربوط به نشانی‌دهی IPv4 ـ ICMP (پروتکل پیام کنترل اینترنت)

۲) قالب دیتاگرام IPv4 - قواعد اداره کردن بسته‌ها در روترها (مثلاً تکه‌سازی و دوباره سرهم کردن) - قواعد مربوط به نشانی‌دهی IPv4

۳) قالب دیتاگرام IPv4 ـ پروتکل‌ها و الگوریتم‌های مسیریابی (مثل OSPF و BGP) ـ قواعد مربوط به نشانی‌دهی IPv4

۴) قالب دیتاگرام IPv4 ـ پروتکل‌های کنترلر SDN ـ قواعد مربوط به نشانی‌دهی IPv4

۱۸- **فرض کنید در روش امضای دیجیتال ElGamal، امضاکننده (آلیس) از k یکسان برای امضای دو پیام مختلف استفاده می‌کند، در این‌صورت کدام‌یک از گزینه‌ها نادرست است؟ (k عدد تصادفی است که توسط آلیس برای امضای پیام انتخاب می‌شود.)**

۱) مهاجم قادر است k را به‌دست آورد.

۲) حمله Existential forgery امکان‌پذیر می‌شود.

۳) مهاجم قادر است کلید خصوصی آلیس را به‌دست آورد.

۴) مهاجم قادر است امضای آلیس را روی یک پیام دیگر جعل کند.

۱۹- کدام‌یک از سرویس‌های امنیتی توسط پروتکل AH (Authentication Header) تامین نمی‌شود؟

۱) جلوگیری از حمله پیام‌های تکرار ۲) احراز اصالت پیام

۳) کنترل دسترسی ۴) محرمانگی

۲۰- در الگوریتم توافق کلید Diffie-Hellman، کدام‌یک از گزاره‌ها در مورد حمله Man in the Middle (مردی در میانه) درست است؟

۱) حتی اگر هر دو طرف گواهی دیجیتالی داشته باشند (بدین معنی که سهم‌های کلیدشان، امضای شخص ثالث قابل اعتماد را داشته باشد)، حمله مردی در میانه قابل انجام است، زیرا این الگوریتم ذاتاً در برابر این حمله امن نیست.

۲) اگر یکی از دو طرف گواهی دیجیتالی داشته باشد (بدین معنی که سهم کلیدش در الگوریتم، امضای شخص ثالث قابل اعتماد را داشته باشد)، حمله مردی در میانه کاملاً غیرقابل انجام می‌شود.

۳) حمله مردی در میانه تنها در لحظه تبادل سهم‌های کلید دو طرف روی کانال قابل انجام است و پس از آن امکان لو رفتن کلید و انجام حمله وجود ندارد.

۴) انجام حمله مردی در میانه حتی پس از پایان مرحله تبادل سهم‌های کلید دو طرف هم ممکن است و امکان بازیابی کلید توسط دشمن وجود دارد.

۲۱- در الگوریتم رمزنگاری 3DES، سه واحد رمزنگاری به‌صورت سری پشت هم قرار داده می‌شوند. رابطه چنین رمزکننده‌ای به‌صورت $C = E(K_3, D(K_2, E(K_1, P)))$ است که P همان Plain Text و C همان Cipher Text بوده و در آن اپراتور $E(K, \circ)$ به معنی رمز کردن با کلید K و $D(K, \circ)$ به معنای رمزگشایی با کلید K است. رابطه رمزگشای 3DES نیز به‌طور مشابه قابل نوشتن است. طول کلید مؤثر در الگوریتم 3DES چقدر است و دلیل استفاده از واحد میانی رمزگشا (D) در میان دو رمزکننده (E) در طراحی این رمزکننده چیست؟

۱) طول کلید مؤثر ۱۶۸ بیت است. استفاده از D در مرحله میانی باعث امنیت بیشتر می‌شود و شکستن الگوریتم را مشکل‌تر می‌کند.

۲) طول کلید مؤثر ۱۶۸ بیت است. استفاده از D در مرحله میانی به کاربران 3DES امکان رمزگشایی پیام‌هایی که از سیستم‌های قدیمی‌تر DES آمده است را با تنظیم کلیدهای مناسب، می‌دهد.

۳) طول کلید مؤثر ۱۱۲ بیت است، زیرا کلید K_1 و K_3 باید حتماً یکسان انتخاب شوند. استفاده از D در مرحله میانی باعث امنیت بیشتر می‌شود و شکستن الگوریتم را مشکل‌تر می‌کند.

۴) طول کلید مؤثر ۱۱۲ بیت است، زیرا کلید K_1 و K_3 باید حتماً یکسان انتخاب شوند. استفاده از D در مرحله میانی به کاربران 3DES امکان رمزگشایی پیام‌هایی که از سیستم‌های قدیمی‌تر DES آمده است را با تنظیم کلیدهای مناسب، می‌دهد.

۲۲- در یک سیستم RSA، کلید خصوصی $d = ۷$ و مقدار (تابع فی اولر پیمانه) $\phi(n) = ۷۲$ است. اگر $p = ۷$ یکی از اعداد اول سازنده پیمانه باشد، مقادیر کلید عمومی (e) و پیمانه (n) کدام هستند؟

۱) $e = ۳۱, n = ۹۱$

۲) $e = ۱۷, n = ۹۱$

۳) $e = ۳۱, n = ۷۷$

۴) $e = ۱۷, n = ۷۷$

۲۳- در پایگاه‌های داده توزیع‌شده و مبحث پروتکل نهایی (Commit) تراکنش در دو فاز، اگر سایت هماهنگ‌کننده در هنگام اجرای این پروتکل برای تراکنش T دچار خرابی شود، سایر سایت‌های فعال (مشارکت‌کننده) باید در مورد سرنوشت T تصمیم بگیرند. چند مورد از عبارت زیر در این خصوص درست است؟

- اگر یک سایت فعال حاوی یک رکورد < commit T > در فایل گزارش (Log) خود باشد، T باید تأیید نهایی شود.
- اگر یک سایت فعال حاوی یک رکورد < abort T > در فایل گزارش خود باشد، T باید لغو شود.
- اگر حتی یک سایت فعال حاوی یک رکورد < ready T > در گزارش خود نباشد، سایت هماهنگ‌کننده مربوطه نمی‌تواند تصمیم به تأیید T گرفته باشد، بنابراین T باید لغو شود.
- در حالت غیر از موارد فوق، تمام سایت‌های فعال دارای یک رکورد < ready T > در گزارش‌های خود بوده، اما هیچ رکورد کنترلی دیگری در مورد T وجود ندارد. بنابراین از آنجایی‌که هماهنگ‌کننده هم دچار خرابی شده است، T باید لغو شود.

۱) ۱ ۲) ۲ ۳) ۳ ۴) ۴

۲۴- وظیفهٔ تأمین خاصیت دوام (Durability) بر عهده کدام زیرسیستم است؟

۱) Backup Subsystem
۲) Recovery Subsystem
۳) Concurrency Control Subsystem
۴) این خاصیت مرتبط با هیچ زیرسیستمی نیست و مرتبط با رسانه ذخیره‌سازی، یعنی حافظه‌های دائمی مانند هارددیسک است.

۲۵- نوع دیگری از پروتکل درخت (Tree Protocol) وجود دارد که نام آن پروتکل جنگل (Forest Protocol) است. پایگاه داده در پروتکل جنگل به شکل یک جنگل از درخت‌های ریشه‌دار است. تمامی قوانین این دو پروتکل یکسان هستند. به‌جز یک قانون که هر تراکنش T_i، اولین قفل بر روی هر یک از درخت‌ها را می‌تواند بر روی هر گره دلخواه اعمال کند. اکنون طبق پروتکل جنگل کدام مورد درست است؟

۱) این پروتکل گرسنگی (Starvation) ندارد ولی عدم وجود بن‌بست (Deadlock) را تضمین می‌کند.
۲) این پروتکل توالی‌پذیری نمایی (View Serializable) و همچنین توالی‌پذیری نتیجه‌ای را تضمین نمی‌کند.
۳) این پروتکل قابلیت بازیابی (Recoverable) ندارد ولی توالی‌پذیری تعارضی (Conflict Serialization) را تضمین می‌کند.
۴) این پروتکل توالی‌پذیری تعارضی (Conflict Serialization) و همچنین طرد تسلسلی (Cascadeless) را تضمین می‌کند.

۲۶- یک مدیر کنترل همروندی مبتنی بر مهر زمان سختگیرانه (Strict Timestamp Ordering) را در نظر بگیرید. در زیر دنباله‌ای از رویدادها شامل رویدادهای شروع، که در آن St_i یعنی تراکنش T_i شروع می‌شود و co_i یعنی تراکنش T_i کامیت می‌شود. این دنباله نشان‌دهنده زمان واقعی است و زمان‌بند مبتنی بر مهر زمان به تراکنش‌ها، مهر زمان را به‌ترتیب شروع‌شان تخصیص می‌دهد. مشخص کنید که برای آخرین دستور چه اتفاقی می‌افتد؟ (ترتیب دستورات دنباله از چپ به راست است.)

$St_1 ; St_2 ; St_3 ; r_1(A) ; w_1(A) ; r_2(A)$

۱) بازگشت داده می‌شود. (rolled back)
۲) صرف‌نظر می‌شود. (ignored)
۳) به تأخیر می‌افتد. (delayed)
۴) اجرا می‌شود. (accepted)

۲۷- رابطه $R(A,B,C,D)$ را درنظر بگیرید که در آن A کلید اصلی رابطه است. فرض کنید R_1، R_2 و R_3 سه قطعه (fragmentation) از رابطه R باشند که به‌صورت زیر تعریف شده‌اند. کدام‌یک از معیارهای صحت (Correctness) نقض شده است؟

$R_1 = \pi_{AB}\sigma_{A\geq 2}(R)$

$R_2 = \sigma_{A<2}\pi_{AB}(R)$

$R_3 = \pi_{CD}(R)$

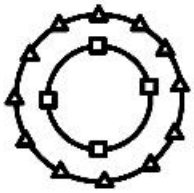
۱) بازسازی (Reconstruction)
۲) جدایی (Disjointness)
۳) کامل بودن (Completeness)
۴) هم‌پوشانی (Overlap)

۲۸- تعداد تکرار کلمات در اسناد D_1 و D_2 به شرح جدول زیر است که در آن T_iها ترم‌های موجود در اسناد هستند. عبارت $T_1T_3T_6$ به عنوان پرس‌وجوی ورودی دریافت می‌شود. درصورتی‌که از مدل زبانی یونیگرم برای بازیابی استفاده شود، امتیاز query likelihood برای دو سند D_1 و D_2 به‌ترتیب (از راست به چپ) چند خواهد شد؟

	T_1	T_2	T_3	T_4	T_5	T_6
D_1	۵	۰	۵	۳	۲	۱۰
D_2	۱	۷	۵	۴	۳	۰

۱) ۰/۰۸ و ۰/۰۳
۲) ۰/۱۶ و ۰
۳) ۰/۸ و ۰/۳
۴) ۰/۰۱۶ و ۰

۲۹- در یک سیستم بازیابی، سندها را با بردارهای دوبعدی نمایش می‌دهند. برای پرس‌وجوی P مجموعه‌ای به‌صورت زیر گزارش شده است که در آن اسناد مرتبط (مربع) و غیرمرتبط(مثلث) روی دو دایره متحدالمرکز با فاصله مساوی قرار دارند. الگوریتم فیدبک روکیو (Rocchio's feedback algorithm) با ضرایب اهمیت یکسان اسناد مرتبط و غیرمرتبط، پرس‌وجوی جدید Q را از روی P می‌سازد. درباره این دو پرس‌وجو کدام گزینه درست است؟



۱) به‌دلیل توزیع متوازن اسناد پاسخ، برای Q دقیقاً همان مجموعه جواب P بازیابی می‌شود.
۲) به‌دلیل توزیع هم مرکز اسناد، برای Q اسناد مرتبطِ کاملاً متفاوتی بازیابی می‌شود.
۳) به‌دلیل تعداد بیشتر اسناد غیرمرتبط، برای Q اسناد مرتبط دقیق‌تری بازیابی می‌شود.
۴) به‌دلیل تراکم بیشتر اسناد مرتبط، برای Q اسناد مرتبط بیشتری بازیابی می‌شود.

۳۰- فرمول هموارسازی Absolute discounting(AD) و Dirichlet Prior(DP) و Jelinek-Mercer(JM) برای محاسبه احتمال وقوع ترم w در سند d را در نظر بگیرید. فرض کنید c(w,d) تعداد تکرار w در d، $|d|_u$ تعداد ترم‌های یکتای d، $p(w|C)$ احتمال وقوع w در مجموعه مرجع C و $|d|$ طول سند d، δ پارامتر AD، μ پارامتر DP و λ پارامتر JM هستند. حال فرض کنید فرمول کلی ما برای هموارسازی به صورت زیر است. در این صورت مقدار x در روش AD، DP و JM به ترتیب کدام است؟

$$P(w|d)=\begin{cases} p_{seen}(w|d) & \text{اگر w در d وجود دارد:} \\ xp(w|c) & \text{در غیر این صورت:} \end{cases}$$

۱) $\frac{|d|_u}{\delta|d|}$ و $\frac{|d|+\mu}{\mu}$ و λ

۲) $\frac{\delta|d|_u}{|d|}$ و $\frac{\mu}{|d|+\mu}$ و λ

۳) $\frac{|d|_u}{|d|}$ و $\frac{|d|+\mu}{\mu}$ و $\frac{\lambda}{|d|}$

۴) $\frac{\delta|d|_u}{\delta|d|}$ و $\frac{\mu}{|d|+\mu}$ و $\frac{\lambda}{|d|}$

۳۱- چه تعداد از گزاره های زیر درباره الگوریتم LSI (Latent Semantic Indexing) درست است؟

- با کاهش تعداد ابعاد، فراخوانی (Recall) بهبود می‌یابد.
- ابعاد به‌دست آمده از این الگوریتم شبیه خوشه‌های معنایی کلمات هستند.
- هزینه محاسباتی تجزیه ماتریس یکی از چالش‌های اصلی این الگوریتم است.
- الگوریتم LSI به‌صورت اتوماتیک تعداد ابعاد بهینه برای نمایش اسناد را می‌یابد.
- به کمک مقادیر منفی در ابعاد حاصله، می‌توان پرس‌وجوهای منفی را هم پاسخ داد.
- کلمات کم بسامد یا دارای خطای املایی سبب ایجاد مقادیر منفی در ابعاد نهایی اسناد هستند.

۱) ۵ ۲) ۴ ۳) ۳ ۴) ۲

۳۲- تعداد خوشه‌بندی‌های ممکن سلسله مراتبی N نمونه براساس شباهت دوبه‌دوی آنها با تعداد کدام‌یک از مجموعه‌های زیر هم‌رتبه است؟

۱) تعداد درخت‌های دودویی دارای N گره

۲) تعداد جایگشت‌های غیرتکراری N شی متمایز

۳) تعداد زیرمجموعه‌های یک مجموعه دارای N عضو

۴) تعداد درخت‌های پوشای کمینه‌گرافی با N رأس و $M > N$ یال

۳۳- یک frequent itemset به‌نام L و یک زیرمجموعه غیرتهی از آن به‌نام S را در نظر بگیرید. همچنین فرض کنید که S_1 زیرمجموعه‌ای از S است. حال دو قانون زیر را در نظر بگیرید. کدام جمله درست است؟

$S \rightarrow (L-S)$

$S_1 \rightarrow (L-S_1)$

۱) مقدار Support قانون $S \rightarrow (L-S)$ بزرگتر از قانون $S_1 \rightarrow (L-S_1)$ است.

۲) مقدار Confidence قانون $S_1 \rightarrow (L-S_1)$ بزرگتر مساوی قانون $S \rightarrow (L-S)$ است.

۳) مقدار Confidence قانون $S \rightarrow (L-S)$ بزرگتر مساوی قانون $S_1 \rightarrow (L-S_1)$ است.

۴) نمی‌توان دقیق اظهارنظر کرد.

۳۴- در مورد گزاره‌های زیر کدام گزینه درست است؟

گزارهٔ A: مجموعه داده X دارای ۱۰۰ بعد است. با استفاده از PCA ابعاد آن به ۲۰ بُعد کاهش می‌یابد. سپس مجدداً روی این مجموعهٔ جدید الگوریتم PCA اعمال شده و ابعاد آن به ۳ بُعد کاهش می‌یابد.

گزارهٔ B: مجموعه داده X دارای ۱۰۰ بُعد است. با استفاده از PCA ابعاد آن به ۳ بُعد کاهش می‌یابد.

۱) لزوماً ابعاد به‌دست آمده در هر گزاره مشابه نیستند.

۲) ابعاد به‌دست آمده در هر دو گزاره مشابه هم هستند.

۳) ابعاد به‌دست آمده در هر دو گزاره مشابه نیستند.

۴) به‌دلیل حذف همبستگی ویژگی‌ها در مرحلهٔ اول، مرحلهٔ دوم گزاره A انجام‌پذیر نیست.

۳۵- اگر برای تشخیص ناهنجاری از الگوریتم Isolation Forest استفاده شود، داده‌های ناهنجار معمولاً چه جایگاهی در درخت‌های تصمیم به‌دست آمده پیدا می‌کنند؟

۱) به‌دلیل تفاوت عمده حداقل یکی از ویژگی‌های آنها با داده‌های عادی، در مسیرهای نزدیک به ریشه طبقه‌بندی می‌شوند.

۲) به‌دلیل شباهت زیاد نمونه‌های ناهنجار به همدیگر، در یک زیردرخت جداگانه طبقه‌بندی می‌شوند.

۳) به‌دلیل شباهت زیاد آنها به نمونه‌های عادی، در همه جای درخت طبقه‌بندی می‌شوند.

۴) به‌دلیل تعداد کم آنها در مسیرهای طولانی و دور از ریشه طبقه‌بندی می‌شوند.

۳۶- تجارت الکترونیکی در دورهٔ گذار از محیط‌های سنتی به محیط‌های مبتنی بر I.T. منجر به کدام مورد می‌شود؟

۱) افزایش قدرت چانه‌زنی مشتریان
۲) کاهش تهدید کالای جایگزین
۳) افزایش قدرت کانال‌های توزیع کالا
۴) افزایش موانع ورود تازه‌واردها به صنعت

۳۷- کدام گزینه، زیرمجموعهٔ تبلیغات مستقیم از طریق موتورهای جستجو به‌شمار می‌آید؟

۱) تبلیغات کلیکی ـ تبلیغات ارگانیک

۲) تبلیغات پولی ـ تبلیغات ارگانیک

۳) تبلیغات کلمات کلیدی ـ تبلیغات کلیکی

۴) افزایش رتبه وب‌سایت (SEO) ـ تبلیغات کلمات کلیدی

۳۸- نسبت مشتریانی که واقعاً از یک وب‌سایت خرید می‌کنند به کل کاربران بازدیدکننده از آن وب سایت، چه نام دارد؟

۱) جذابیت
۲) نرخ کلیک
۳) نرخ همسان
۴) نرخ تبدیل

۳۹- **کدام گزاره درخصوص سرمایه‌پذیری جمعی (crowd-funding) درست است؟**

۱) سرمایه‌پذیری جمعی محدود به رمز ارزها است.

۲) سرمایه‌پذیری جمعی نوع خاصی از سرمایه‌گذاری فرشته (angel) است.

۳) در سرمایه‌گذاری جمعی تمامی قوانین رایج و متعارف سرمایه‌گذاری در بازار سرمایه رعایت می‌شود.

۴) ریسک‌های این نوع سرمایه‌گذاری نسبت به دیگر سرمایه‌گذاری‌ها نظیر بازار بورس برای سرمایه‌گذاران بالاتر است.

۴۰- **کد دیجیتال گنجانده شده در یک تصویر دیجیتال یا فایل صوتی به‌طوری که قابل رؤیت نباشد، چه نامیده می‌شود؟**

۱) علامت تجاری یا تریدمارک ۲) پنهان‌نگاری یا استگانوگرافی

۳) واترمارک ۴) رمزنگاری

۴۱- **کدام مورد تعریف دقیق‌تری از ابزارهای تعاملی ارائه می‌دهد؟**

۱) ابزارهایی که به دانشجویان امکان می‌دهد تا با یک یا چند دانشجوی دیگر تعامل برقرار کنند و به‌صورت مشارکتی بیاموزند.

۲) ابزارهایی که برای کاربردهایی نظیر ویدئو کنفرانس و استریم و پخش زنده مورد استفاده قرار می‌گیرند.

۳) ابزارهایی که برای برقراری ارتباط بین استاد و دانشجویان به‌طور فعال مورد استفاده قرار می‌گیرد.

۴) فرایندهایی که برای برقراری ارتباط بین دانشجویان در یک کلاس مورد استفاده قرار می‌گیرد.

۴۲- **کدام گزاره درخصوص یادگیری مشارکتی درست است؟**

۱) یادگیری مشارکتی می‌تواند هم حالت هم‌زمان (سنکرون) و هم غیرهم‌زمان (غیرسنکرون) باشد.

۲) یادگیری مشارکتی از اولین نسل سیستم‌های یادگیری الکترونیکی مد نظر قرار گرفت.

۳) استفاده از شبکه‌های اجتماعی، آسیب جدی یادگیری مشارکتی به‌شمار می‌آید.

۴) قوت یادگیری مشارکتی بیشتر در آموزش مسائل پیچیده‌تر آشکار می‌شود.

۴۳- **کدام مورد دربارهٔ آموزش مجازی و یادگیری آنلاین درست نیست؟**

۱) آموزش مجازی، سرعت و زمان یادگیری را با یادگیرنده تطبیق می‌دهد.

۲) آموزش مجازی دسترسی به دوره‌های درسی بیشتری را برای یادگیرنده فراهم می‌کند.

۳) در آموزش مجازی، فضای بهتری برای بحث و تبادل نظر میان معلم و یادگیرنده برقرار است.

۴) در آموزش مجازی معلم مجبور نیست که برای جلب توجه دانشجو تلاش کند و فقط باید محتوای خوبی تهیه کند.

۴۴- **آموزش مجازی سنکرون یا هم‌زمان به چه معنا است؟**

۱) دانشجویان با به‌کارگیری سیستم مدیریت یادگیری به محتوای درسی دسترسی پیدا می‌کنند و در مورد آن نظر می‌دهند.

۲) دانشجویان با به‌کارگیری سیستم مدیریت یادگیری یا هر سامانه دیگری از طریق شرکت در کلاس مجازی، وبینار و یا تلفن به یادگیری می‌پردازند.

۳) دانشجویان از طریق تالارهای گفتگو دربارهٔ مطالب درسی به بحث و تبادل نظر می‌پردازند.

۴) دانشجویان در زمان و مکان دلخواه خود به یادگیری می‌پردازند.

۴۵- **کلاس معکوس یا Flipped-Classroom توسط کدام عبارت بهتر تبیین و تشریح می‌شود؟**

۱) معلم مطالب را از قبل آماده کرده و در اختیار دانشجو قرار می‌دهد. در کلاس دانشجو با آمادگی مطالب را با هم‌کلاسی‌ها مباحثه کرده و نقش معلم هماهنگی است.

۲) معلم به روش سنتی مطالب را آماده می‌کند ولی به دانشجویان تکلیف داده می‌شود که آنها هم مطالب را دوباره تهیه کنند.

۳) تنها تکلیف دانشجویان مشارکت در بحث است و خارج از کلاس کاری ندارند.

۴) تمرکز در این روش بر یادگیرنده است و معلم نقش خاصی ندارد.