

מילון מונחים

(לפי סדר הא'-ב')



אלגוריתם (Algorithm) – מכלול של הוראות או הנחיות המבוצעות שלב-אחר-שלב כדי לבצע משימה מסוימת או לפתור בעיות. מנועי חיפוש למשל, משתמשים באלגוריתמים שונים כדי להציע למשתמשים את התוצאות הרלוונטיות ביותר עבורם.

אלגוריתם דרוג (Ranking Algorithm) – אלגוריתם המדרג את תוצאות החיפוש במנועי החיפוש לפי הרלוונטיות שלהם לשאלת החיפוש.

ארגונים המונעים ממידע (Data Driven Organization) – ארגונים העושים שימוש נבון במערכות ביג דאטה לפעילותם השוטפת.

ביג-דאטה (Big Data) – נתוני עתק, מידע בהיקפים גדולים מאוד ממקורות מידע שונים ומגוונים המאוחסן לאחר שעבר תהליך דיגיטציה המאפשר אחזורו במהירות, באופן שמאפשר ליצור ערך משמעותי ולהצביע על מתאמים בין מרכיבי המידע. נהוג לטעון שפרדיגמת הביג דאטה שונה מפרדיגמות מידע קודמות בשלושה מאפיינים, שמכונים שלושת ה-V: נפח המידע (Volume), מגוון המידע (Variety) והמהירות שבה יש לטפל במידע (Velocity).

בינה מלאכותית (Artificial Intelligence) – היכולת של מחשב לבצע משימות המדמות את האינטליגנציה האנושית, כמו היכולת להסיק מסקנות, ללמוד מניסיון עבר וכדומה.

"גוגל סג'סט" (Google Suggest) – פונקציה קיימת בגוגל, המשלימה באופן אוטומטי שאילתות חיפוש. הפרויקט מחייב כמויות עצומות של מידע (מיליוני משתמשים), ופלטפורמות חישוב מקביליות (MapReduce).

דיבור לטקסט (Speech to text) – טכנולוגיה שמטרתה להפוך שמע לטקסט, ככלי המסייע בהתמודדות עם מאגרי מידע עצומים שאינם מובנים.

האינטרנט של הדברים (IoT, Internet of Things) – מונח זה מתאר חפצים שניתן לתקשר איתם ושיכולים לתקשר ביניהם באופן ממוחשב. תפיסה זו מאפשרת, בין היתר, יצירתם של "בתים חכמים" ו"ערים חכמות".

ווב סמנטי (Semantic Web) – יכולת לקשור בין נתונים בבסיסי הנתונים, על ידי ייצוג אחיד של המידע, זאת תוך יצירת קשרים בין ישויות על פי התכונות המשותפות שלהן, על פי הקשרים שלהן עם ישויות אחרות, וכן על פי מרכיבים רבים אחרים שהמשתמש מסוגל להגדיר.

חוכמת ההמונים (Crowdsourced intelligence) – מודל מבוסס טכנולוגיות מידע, המיועד לפתרון בעיות ולפיתוח רעיונות, הממנף ידע מבוזר שנמצא בידי קבוצות ויחידים, במטרה לייצר משאבים מגוונים לארגונים. בהקשרי מודיעין, חוכמת ההמונים עשויה לסייע באיסוף מידע, בפיתוח ידע ובחיזוי וניבוי. המאמר העוסק בכך טוען כי חוכמת ההמונים היא דיסציפלינה ייחודית, השונה מהדיסציפלינות האחרות, ומשך מציע לכנותה בשם: CROSINT (בדומה ל-OSINT, HUMINT וכו').

טכנולוגיות האצה חומרית – טכנולוגיה של מעבדים גרפיים, ששימשה בעבר לשיפור איכות

התמונה במשחקי מחשב, נמצאה מתאימה מאוד להרצה של אלגוריתמים מקביליים של ביג דאטה. תחום הביג דאטה הפך ליעד מסחרי מרכזי של חברות שמייצרות חומרה כזו.

טכנולוגיות עיבוד מבוזר – טכנולוגיות המאפשרות את הבאת המידע אל יחידת העיבוד. נדרשות נוכח הימצאותם של מאגרי מידע אדירים בידי ענקיות הטכנולוגיה הגדולות (Google, Facebook, Twitter, Amazon) באופן מבוזר.

טקסט כדאטה – כלי מחקרי אשר נעשה בו שימוש גובר במחקרי ביטחון ומדיניות, ובבסיסו ההנחה כי כדי לזהות, לאפיין ולהעריך את משמעות הדפוסים והמגמות בשיח, יש לייחס חשיבות רבה לאופן שבו שחקנים מייצרים שיח פומבי, המתקף ומתרגם את האינטרסים שלהם לפרקטיקות התנהגותיות.

כריית מידע (Data Mining) – תהליך המאפשר להפוך מידע גולמי לידע שימושי, תוך שימוש בכלים שונים המאפשרים לזהות חריגות משגרה, תבניות ומתאמים במאגרי מידע גדולים. בשל כך, כריית המידע תלויה באיסוף המידע, באחסונו ובעיבודו.

למידת מכונה (Machine Learning) - תחום זה נחלק למספר תחומים נוספים:

1. **למידה מכוונת (Supervised Learning)** – טעינת מידע מתויג למערכת, כך שהיא לומדת את מאפייניו, ולאחר מכן תוכל לזהות את האובייקט המבוקש במידע חדש שייקלט.

2. **למידה לא מכוונת (Unsupervised Learning)** – מטרת האלגוריתמים של טכניקה זו היא למצוא ייצוג פשוט וקל להבנה של אוסף הנתונים, ללא תיוגו מלכתחילה.

3. **למידה מחוזקת (Reinforcement Learning)** – המערכת לומדת מתוך חיכוך ומתוך ניסוי וטעייה.

4. **למידת עומק (Deep Learning)** – אחד מסוגי למידת המכונה, אשר אינה מחייבת בחירה אנושית של מאפייני המידע מראש. סוג זה של למידת מכונה משתמש ב"רשתות נוירונים מלאכותיות" (המדמות את הפעילות החישובית של רשתות נוירונים במוח האנושי) בעלות מספר רב של שכבות ביניים.

מאגרי המידע (Dataset):

1. **GDELТ (The Global Database of Events' Language and Tone)** – מאגר שפותח על ידי אוניברסיטת ג'ורג'טאון, שסורק, מקודד ומנגיש לשימושים ציבוריים ומחקריים 98% מכל התקשורת העולמית.

2. **ICEWS (The Integrated Crisis Early Warning System)** – כלי לחיזוי משברים שפותח על ידי המערכת הביטחונית האמריקאית שאוסף ומקודד נתוני תקשורת גלויה כמצע מאפשר לחיזוי והתרעה על משברים. הכלי פותח בעיקר בראייה ביטחונית כדי לסייע בבניין הכוח, וכיום נטען כי השימוש בו מאפשר לחוקר להגיע לדיוק של 80% בחיזוי והתרעה על משברים (פונקציית החיזוי אינה נגישה לציבור הרחב ולאנשי האקדמיה).

מדור מידע וקמ"נות מידע – תפיסה לפיה יחידות המודיעין נדרשות לקמ"ני מידע משלהן, בעלי זיקה טכנולוגית, לשיפור יכולות המיצוי וההפקה של היחידות ע"י מיצוי מאגרי מידע קיימים; פיתוח נגישויות והקמת מאגרים חדשים; ועיבוד המידע הקיים לכדי תוצר מודיעיני משודרג. מדורים אלה ברובם אחראים גם על הסיוע בהכשרות החוקרים, על ידי שיפור מיומנויות החוקר לטפל במידע והנגשת מערכות מידע קיימות.

מהפכה בעניינים מודיעיניים (Revolution in Intelligence Affairs) – שינוי תפיסתי וארגוני משמעותי במערכת המודיעין כדי להתאים את הפרדיגמה המודיעינית הקיימת לצרכיה העכשוויים של קהילת המודיעין, בשל השתנות הכישורים, ההתארגנות או התהליכים הנדרשים, ובשל השתנות משימותיה.

מחקר רשתי (Social Network Analysis, SNA) – זיהוי מאפייניה של רשת חברתית באמצעות הסתכלות רחבה ככל שניתן על הרשת ושימוש באלגוריתמיקה המזהה את תכונותיה, ובראשן מיפוי מרכזי הכובד (data-driven). זאת, בשונה מגישת המחקר המתמקדת בניית קשרים (link analysis), שלפיה מבוצע מחקר נקודתי של העוגנים והצמתים מסביבם, המתבצע בהיעדרן של יכולות ביג-דאטה.

1. **רכזות (Hubs)** – צמתים "פופולריים" מאוד המשפיעים על הרשת כולה. הרכזות קיימות כמעט בכל רשת.

2. **רשת מתקהלת** – המבנה הפנימי של הרשת מלמד כי הרשת אינה אקראית, אלא בנויה מצבירים (קהילות), בעלי היגיון של דומות או ריבוי אינטראקציות פנימיות. חלוקה זו יכולה להמחיש את מבנה הארגון "האמיתי".

מחשוב ענן (Cloud Computing) – נסמך על משאבים מחשוביים משותפים, לעומת שרתים מקומיים. הגישה למידע המאוחסן בענן מתבצעת דרך האינטרנט.

מידע מובנה ומידע לא מובנה (Structured and Unstructured Data) – מידע מובנה הוא מידע בעל רמות ארגון וסידור גבוהה, כמו במסדי נתונים רלציוניים (RDBMS). מידע לא מובנה (שמע, תמונה, וידיאו) הוא מידע קשה לאינדוקס, שליפה וחיפוש. מידע טקסטואלי הנכתב באופן חופשי למשל, ע"י איש המודיעין נחשב כמידע לא מובנה, ומידע גיאוגרפי וקטורי הוא למשל מידע מובנה.

1. **אחזור סמנטי** (Semantic Retrieval) – צריכה של מידע לא לפי צורתו הפיזית במאגרי המידע אלא בראי המשמעות המודיעינית מגולמת בו. לפי גישה זו, כדי לזהות חתימה מסוימת במידע על אודות התקשרויות, לא נרצה שאיש המודיעין יהיה אחראי לדעת היכן מפורז המידע מסוג זה, אלא נרצה שיוכל לשאול, והתשתית המודיעינית-טכנולוגית תדע לתווך את הבקשה שלו לכלל המידע הרלוונטי.

2. **אינדוקס** – מלשון אינדקס (Index). לכל חלקיק מידע שנכנס לבסיס הנתונים, יש כתובת שבה הוא "נמצא", וידעתה מאפשרת לשולפו על ידי שאילתה שמוגדרת על ידי חוקר או על ידי מערכת אוטומטית.

מעגל ה-Hype (Hype Cycle) – זוהי הדרך שבה הקהילה הטכנולוגית מאמצת טכנולוגיה חדשה, הכוללת מספר שלבים: אימוץ מהיר (מדי) של פתרון חדש המופיע בשוק; התפכחות – אלה שהפתרון אינו מתאים להם, נוטשים אותו; הארה – מוצעות דרכים נוספות להשתמש בפתרון החדש ועל-ידי כך מורחב מספר הלקוחות שלהם הוא מתאים; ישורת יצרנית – הגעת הפתרון לבגרות מבחינת יכולות ומבחינת האימוץ שלו.

מערכת דירוג חברתית (Social Credit System, SCS) – המשטר הסיני החל לפתח בשנת 2014 את המערכת, ובשנת 2020 צפוי לחייב את כלל האזרחים להשתמש בה. המערכת תנתח את התנהלותם היומיומית של האזרחים, תוך שימוש בטכנולוגיות מתקדמות לכריית מידע ולניתוחו, ועל-בסיס מידע זה יתוגמלו או ייענשו אזרחי המדינה על מעשיהם.

מרכזי היתוך (Fusion Centers) – גופים המאגדים תחת ארגון אב גורמים מכלל הארגונים הרלבנטיים ליצירת תמונת מודיעין (ולעתים גם מבצעים) אחודה, ומטרתם לבחון איזונים מסוימים בהסתכלות על-זירתית ועל-ארגונית, תוך שיתופי פעולה עם כלל הגופים הרלוונטיים (גופי מודיעין, רשויות אכיפת חוק וכו').

פרויקט ה-Hadoop – פרויקט קוד פתוח בשימוש נפוץ בתחום ה-Big Data שמממש תשתית תוכנה לעיבוד מבוזר המתאימה לעיבוד כמויות גדולות של נתונים.

פרשת סנודן – אדוארד סנודן היה עובד חיצוני של ה-NSA. במהלך עבודתו בקהילת המודיעין האמריקנית אסף מסמכים מסווגים, ובשנת 2013 החל להדליף. המסמכים הרבים שהדליף חשפו את מימדי איסוף המידע של קהילת המודיעין האמריקנית ועוררו שיח ציבורי נרחב בנוגע להפרת זכויות אדם ולצורך בעידוד השקיפות.

פרשת קיימברידג' אנליטיקה (Cambridge Analytica) – קיימברידג' אנליטיקה הינה חברת ייעוץ, המשלבת כריית מידע וניתוחו. במהלך מערכת הבחירות בארה"ב בשנת 2016 יצרה פרופיל פסיכולוגי של מצביעים אמריקניים, שנועד לסייע בהכוונת הקמפיינים הפוליטיים. לאחרונה, נחשף כי במסגרת פעילותה עשתה החברה שימוש במידע שאספה מחשבונות פייסבוק של עשרות מיליוני משתמשים.

קוד פתוח (Open-source model) – תוכנה שקוד המקור שלה פתוח ונגיש לכל.

קידוד Cameo (Conflict and Mediation Event Observation) – ספר קידוד המכיל קטגוריות קידוד המאפשרות הבחנה בין קריטריונים רבים (מאות נושאים, אלפי תכונות, אלפי תת-נושאים בדיווחי התקשורת הנאספים במאגר) ומוסכמים, המקובלים זה עשורים אחדים במחקרים במדעי המדינה ומתוקפים במחקרים אקדמיים.

קצירה ממוכנת (Web scrapping) – איסוף והורדת נתונים באופן ממוכן מרשת האינטרנט.

רשתות נוירונים מלאכותיות (Artificial Neural Networks) – רשתות המדמות את הפעילות החישובית של רשתות נוירונים במוח האנושי, ומאפשרות למידת מכונה מקבילית מהירה.

שפת פיית'ון (Python) – שפת תכנות שדרכה ניתן ליישם מניפולציות שונות על דאטה (ברשת ובכלל) ובעיקר לניתוחים סטטיסטיים.

שפת R – שפת תכנות שדרכה ניתן ליישם מניפולציות שונות על דאטה (ברשת ובכלל) ובעיקר לניתוחים סטטיסטיים.

שפת SQL (Structured Query Language) – שפת מחשב המתאפיינת בגמישות גבוהה לתשאול מאגרי מידע ולאחזור מידע מובנה.

תוכנת QDA MINER – מאפשרת ניתוח תוכן באופן ממוחשב. תוכנה זו היא אחת מאפליקציות ניתוח התוכן המובילות והמתקדמות ביותר במרחב המחקר האקדמי.

תפקידים חדשים בעידן המידע:

1. **אזרח מידע** (Data Citizen) – בעל "אוריינות או תודעת מידע". התפיסה היא שחוקרי המודיעין בעתיד צריכים להיות בעלי יכולת זו.

2. **מדען מידע** – (Data Scientist) – איש מקצוע האחראי על איסוף וניתוח מידע בהיקפים גדולים באופן המסייע לארגון לשפר את ביצועיו ולצבור יתרון יחסי על-פני יריביו, תוך הפעלת אלגוריתמים מגוונים.

3. **מהנדס מידע** (Data Engineer) – תפקידו סידור ותקנון החומר, ניקוי, והכנסתו למאגרים בצורה שמאפשרת שליפה.