**Анкета**

Структурно організований набір запитань, кожне з яких логічно пов'язане з основною метою дослідження, отримання інформації відбувається шляхом опитування респондентів. Використовується при проведенні соціологічних, політичних, психологічних, маркетингових досліджень для оцінки поточної ситуації або визначення перспектив розвитку досліджуваних процесів.

**Aнкетування**

У соціології анкетування - це метод опитування, який використовується для складання статичних (одноразове анкетування) або динамічних (При багаторазовому анкетуванні) статистичних уявлень з метою прогнозування дій або подій.

**Mетоди Форсайта**

Форсайт як наукове дослідження базується на застосуванні методів організації структурованих дискусій, в ході яких виробляються загальні уявлення про довгострокові перспективи соціально-економічного та науково-технологічного розвитку. Узгоджене уявлення має на увазі не тільки спільні цілі, а й спільні цінності та мову.

**Cтохастичні моделі**

Стохастичні моделі описують випадкові процеси або ситуації, при цьому мається на увазі, що випадковість тих чи інших явищ виражається в термінах ймовірності. Так само, як і детерміновані, стохастичні моделі бувають дискретні і безперервні.

**Точка біфуркації**

Критичний стан системи, при якому система стає нестійкою щодо флуктуацій і виникає невизначеність: чи стане стан системи хаотичним або вона перейде на новий, більш диференційований і високий рівень впорядкованості.

**Флуктуа́ція або флюктуа́ція**

Випадкове відхилення значення фізичної величини від середнього в певній ділянці простору чи в певний момент часу.

Величини флуктуацій у фізиці зазвичай описуються середньо-квадратичним відхиленням.

Флуктуації відіграють велику роль у різних фізичних процесах. Характерним прикладом фізичного явища, в якому визначальними є флуктуації його характеристик є Броунівський рух.

**Лінгвістична змінна**

Лінгвістична змінна - це множина нечітких змінних, що приймає значення з множини слів або словосполук деякої природної мови. Вона використовується для того, щоб дати словесний опис деякому нечіткому числу, отриманому в результаті деяких операцій. Також поняття лінгвістичної змінної відіграє важливу роль в нечіткому логічному виведенні та в ухваленні рішень на основі наближених міркувань.

**Концепція open data**

Відкриті дані — це концепція, за якою певні дані мають бути вільними для використання та розповсюдження будь-якою особою, за умов дотримання правил атрибуції та/або share-alike ліцензії.

При цьому, під вільністю розуміють умови прийнятні для широкого загалу. Так дані можуть бути доступні за невелику плату, що покриває витрати на їх створення та розповсюдження.

Концепція відкритості даних не нова, але її активне поширення почалось з розвитком інформаційних технологій та Інтернету, зокрема. Треба зазначити, що серед множини відкритих даних окрема увага приділяється відкритим державним даним, як інструменту оцінки та контролю роботи влади та держави, що входить до моделі електронного уряду.

**Формати даних**

**CSV** (від англ. comma-separated values ‘значення, розділені комою’, іноді character-separated values ‘значення, розділені символом’) — файловий формат, котрий є відмежовувальним форматом для представлення табличних даних, у якому поля відокремлюються символом коми та переходу на новий рядок.

Поля, що містять коми, декілька рядків, або лапки (позначаються подвійними лапками), мають обмежуватися з обох боків лапками.

Формат CSV використовується для перенесення даних між базами даних та програмами — редакторами електронних таблиць.

Розши́рювана мо́ва розмі́тки (англ. Extensible Markup Language, скорочено XML) — запропонований консорціумом World Wide Web Consortium (W3C) стандарт побудови мов розмітки ієрархічно структурованих даних для обміну між різними застосунками через Інтернет.

Є спрощеною підмножиною мови розмітки SGML. XML-документ складається із текстових знаків, і придатний до читання людиною.

**JSON** (англ. JavaScript Object Notation, укр. запис об'єктів JavaScript, вимовляється джейсон) — це текстовий формат обміну даними між комп'ютерами.

JSON базується на тексті, може бути прочитаним людиною.

Формат дозволяє описувати об'єкти та інші структури даних. Цей формат головним чином використовується для передачі структурованої інформації через мережу (завдяки процесу, що називають серіалізацією.

**Data sets**

Набір даних — колекція однотипних даних, що застосовується в задачах машинної обробки даних.

Найчастіше набір даних відповідає змісту однієї таблиці бази даних або статистичній матриці даних, де кожна з колонок таблиці містить однорідні значення, а кожен з рядків таблиці відповідає певному члену набору даних.

Наприклад, набір даних про квіти може містити назву різновиду, розміри пелюсток, яскравість забарвлення тощо.

Термін набір даних також використовується при визначенні даних в сукупності тісно пов'язаних таблиць, зображень тощо, що описують результати конкретного експерименту або подію.

Прикладом такого типу є набори даних, зібрані космічними агентствами, що виконують експерименти з приладами на борту космічного зонда, або світлини, передані з космосу.

Окремі набори даних широко використовуються в академічних колах як тестові набори, що підтверджують результати наукових досліджень. Окремі набори даних є відкритими для використання, інші надаються за, звичайно символічну, плату.

**NLP — Natural Language Processing** - загальний напрям інформатики, штучного інтелекту та математичної лінгвістики, що вивчає проблеми комп'ютерного аналізу та синтезу природної мови.

**Бібліотека** — збірка об'єктів чи підпрограм для вирішення близьких за тематикою задач.

**Слово** — це послідовність морфем, об'єднаних за граматичними правилами певної мови і співвідносних з певним елементом позамовної реальності.

**Граф** — це сукупність об'єктів із зв'язками між ними.

**Синтаксичний аналіз** — процес аналізу вхідної послідовності символів, з метою розбору граматичної структури згідно із заданою формальною граматикою.

**Контекстний аналіз** — аналіз частини через ціле, коли одиницями аналізу выступают мовні одиниці, більш широкі, ніж сама одиниця.

**Токенізація** — найперший крок в обробці тексту що полягає в розбитті (розділенні) довгих рядків тексту на більш короткі: абзаци діляться на речення, речення на слова.

**Нормалізація** — серія операцій, в результаті яких текст приводиться до однорідного вигляду: всі слова приводяться до одного регістру, видаляються знаки пунктуації, розшифровуються скорочення, числа приводяться до їх текстового написання і т.д.

**Стемізація** — усунення придатків до кореню, тобто відділення суфікса, префікса, закінчення.

**Лематизація** — операція, що приводить слово до смислової канонічної форми слова (інфінітив для дієслова, називний відмінок однини — для іменників та прикметників).

**Корпус** — набір текстів.