**Онлайн-тренингдин кыскача мазмуну**

**«Ачык маалыматтар: негизги түшүнүктөр, булактар жана форматтары, маалыматтарга жеткиликтүүлүк»**

Тренингде төмөнкүлөр жөнүндө сөз болот,

• Маалыматтар кайдан алынат

• Маалыматтын кандай түрлөрү бар

• Ачык маалыматтар деген эмне

• Маалымат булактары

• Маалымат жок болсо, эмне кылуу керек

• Маалыматтын сапатынын критерийлери

Маалыматтар – анализ жүргүзүү үчүн ыңгайлуу түрдө жазылган жана структураланган ар бөлөк маалыматтардын жыйындысы. Маалыматтын чоң көлөмү ырааттуу түрдө жазылганда, өзүнчө жазууларды да, ошондой эле жыйынды маалыматтарды да кийин оңой талдап чыгууга болот. Ошентип, маалыматтар баштапкы материал катары каралат, анын негизинде кайра иштеп чыккандан кийин маанилүү маалыматты чыгарса болот.

Маалыматтар топтомун талдоо менен маселенин масштабын көрсө болот, изилдөө предмети боюнча жаңы билимге ээ болосуз, гипотезаларды текшере аласыз, маалыматтарды бири-бири менен салыштырып, предметти изилдөө үчүн жаңы көз карашта боло аласыз, визуалдаштырууну түзө аласыз жана маалыматтарга негизделген окуяларды жаза аласыз.

**Маалыматтар кайдан алынат?**

Алар биздин айланабызда дайыма бар. Биз аларды дээрлик ар бир дем алуубуз менен кошо генерациялайбыз. Академиялык изилдөөлөр, мамлекеттик органдардын жана коомдук уюмдардын веб-сайттары, спутниктен алынган сүрөттөр, мобилдик телефонуңуздан А чекитинен В чекитине чейин ары-бери кыймыл тууралуу маалыматтар – мунун баары маалыматтар болуп саналат. Бул биздин ар бирибиз, а биз болсо сегиз миллиард адамдын бирибиз, биздин ар бир кадамыбыз, кыймылыбыз жана аракетибиз. Маалыматтар – бул дүйнөдөгү бардык түзүлүштөр болуп саналат, алар азыр жер бетиндеги адамдардан да көп – бул Нерселердин Интернети деп аталган түзүлүш – ал маалыматтарды туруктуу генерациялап турат. Бул маалыматтардын көлөмүн биз элестете албайбыз, алардын масштабы дайыма өсүп турат.

**Маалыматтар ачыктык деңгээли боюнча:**

* Ачык
* Жабык
* Жарым-жартылай ачык деп бөлүнөт

Ачык маалыматтар

“Ачык маалыматтар каалаган максатта каалаган адам тарабынан эркин колдонулушу, өзгөртүлүшү жана таратылышы мүмкүн.”.[[1]](#footnote-1)

Ачык маалыматтардын эл аралык деңгээлде кабыл алынган бул аныктамасы ачык маалыматтардын маңызын мүмкүн болушунча кыскача жалпылайт. Маалыматтар колдонулган, өзгөртүлгөн жана жайылтылган – алар бул үч критерий боюнча каралат жана ачык деп таанылат. Маалыматтар Интернетте жеткиликтүү болсо, колдонулушу мүмкүн. Алар буга жол берген форматта берилсе, өзгөртүлүшү мүмкүн. Акыр-аягында, ага автордук укук берилбесе, маалыматтарды каалаган адам каалаган максатта тарата алат. Ошентип, үч шарт тең иш жүзүндө техникалык болуп саналат - маалыматтарды жарыялоо, формат жана лицензия анын ачыктыгын аныктайт.

Бул маалыматтарды чогултуу үчүн салык төлөөчү катары жарандар тарабынан төлөнгөн деген позициядан алганда ачык-айкындуулук идеологиясы жана ачык маалыматтардын жүгүртүлүшү башталат, демек, ага жеткиликтүүлүк башка негизги укуктар сыяктуу эле ар бир адамдын укугу болуп саналат.

Жабык маалыматтар

Эреже катары, ачык маалыматтардан айырмаланып эки өзгөчөлүк бар – бул жеке маалыматтар, алар жеке жашоо-турмушту коргоо жөнүндө мыйзам менен корголгон, ошондой эле улуттук коопсуздук маалыматтары, алардын ачыкка чыгышы мамлекеттин өзүнө коркунуч жаратат.

Ошентсе да, алгач жеке маалыматтарды камтыган маалыматтар да ачыкка чыгышы мүмкүн, эгерде аларда айрым адамдар боюнча маалымат анонимдүү болсо, б.а конкреттүү адамдар идентификацияланышы мүмкүн эмес.

Мындай маалыматты берүүнүн дагы бир варианты – топтолгон маалыматтар, башкача айтканда, ар бир адам боюнча маалыматтар эмес, маанисинин суммасы саналат.

Жарым-жартылай ачык маалыматтар

“ Жарым-жартылай ачык маалыматтар” – бул ачык маалыматтардын идеясынын өзү расмий түрдө таанылат, бирок бардык реалдуу иш, анык маалыматтар же жарыяланбайт же машина окубаган формага ээ же анча көрүнүктүү эмес жерде болот. Мисалы, “жарым-ачык маалыматтар” категориясына төмөнкүлөрдү таандык кылса болот. Алар мамлекеттик ведомствонун сайтында жарыяланган, алар жүктөө үчүн жеткиликтүү, бирок машина окуй турган форматта эмес *(мисалы, PD форматында, сүрөттөн түзүлгөн), бул аларды иштетүү боюнча талдоо ишин кыйындатат.*

**Деталдаштыруунун деңгээлине жараша маалыматтар микромаалыматтарга жана топтолгон маалыматтарга бөлүнөт:**

Микромаалыматтар – бул биз бир байкоого чейин көз сала турган маалыматтар. Мисалы, спорттук тиркеменин пайдалануучуларынын кыймылынын траекториясы бизге ар бир пайдалануучунун жолун көрсөтүп турат.

Топтолгон маалыматтарда биз конкреттүү адамдын жолун көрбөйбүз, бирок кайсы шаарларда спорттук тиркеменин пайдалануучулары аз же көп экендигин көрө алабыз.

Көбүнчө микромаалыматтарды колдонуу ыңгайлуу, анткени топтолгон маалыматтарда бизге керектүү деталдаштыруунун деңгээли болбошу да мүмкүн.

**Маалыматтардын форматтары кандай болот?**

Көбүнчө **тексттик маалыматтар** менен иштөөгө туура келет. Бул эң татаал форматтардын бири. Эгерде тексттик документ бирөө болсо, мисалы, чечим бирөө болсо, анда биз аны көзүбүз менен окуп, керектүү маалыматты ошол жерден жазып алабыз. Эгерде мындай чечимдер миңдеген болсо, аларды иштеп чыгуу үчүн код жазууга туура келет.

*Кийинки маалымат форматы* ***PDF*** *(Portable Document Format – электрондук документтердин платформалар аралык ачык форматы).* Ал ошондой эле машина менен иштеп чыгуу үчүн ыңгайлуу эмес. PDF файл документтен түзүлгөн болсо, анда бул жакшы. Биз аны таблицага айландырып, аны менен иштей алабыз. PDF сүрөттөн түзүлгөн болсо, анда бул ыңгайсыз. Кыязы, биз аны тааный албайбыз жана андан маалыматтарды алуу үчүн убакытты текке кетиребиз.

Эң ыңгайлуу жана кеңири таралган форматтардын бири – бул **CSV** же үтүр менен бөлүнгөн маанилер. Бул акысыз, аны менен бөлүшүү жана таблицалык редакторлоруна импорттоо оңой.

Дагы бир формат – бул **JSON.** Ал маалыматтардын кыйла татаал структуралары үчүн арналган. JSON жардамы менен биз маалыматтарды өзүнчө саптарга жаза албайбыз. Эгерде JSON форматында чоң көлөмдөгү маалыматтар сакталса, анда алар менен иштөө үчүн жөнөкөй таблицалык редакторлор жетишсиз. Аны өзгөртүү үчүн сизге кодду жазууга туура келет. JSON файлын тез көрүү үчүн браузерди кеңейтүү бизге жардам берет.

Дагы бир маалымат форматы – бул **XML** (маалыматтар Microsoft Excel таблицаларында сакталат). Бул белги тили, анын жардамы менен маалыматтардын ар кандай касиеттери берилет. Бул формат сейрек кездешет, бирок ошентсе да, маалыматтар ушул форматта да кездешет.

Кандай маалымат булактары бар?

Эң чоң маалымат берүүчү – бул мамлекет болуп саналат. Ал өзүнүн иши жөнүндө маалыматтарды чогултууга милдеттүү. Мындан тышкары, мамлекет кээ бир изилдөөлөргө, мисалы, калкты каттоо сыяктуу ири масштабдуу жана көп ресурстарды талап кылган изилдөөлөргө монополияга ээ.

Мамлекеттик маалыматтардын оң жана терс жактары болот:

Оң жактары төмөнкүлөр болуп саналат:

• Алар көп, кээ бир маалыматтар машина окуй турган форматтарга которулган жана дайыма жаңыланып турат

• Алардын көпчүлүгү акысыз

• Көп учурда бул маалыматтарды мамлекеттен башка эч ким чогулта албайт

Терс жактары төмөнкүлөр болуп саналат:

• Көптөгөн эски жана жаңыланбаган маалымат топтомдору

• Көрсөткүчтөрдү бириктирүү

• Басылманын стандарттарынын жоктугу

• Бирдиктүү иштөөчү издөө механизминин жоктугу

• Ар бир адамда эле API боло бербейт

Мындан тышкары, маалыматтарды бизнес кампаниялар чогултушу жана бериши мүмкүн. Алар өз маалыматтарын сейрек ачыкка чыгарышат. Бирок мындай да болушу мүмкүн. Бизнес-компаниялардын маалыматтары кыйла жогорку сапаттуулукта болот, анткени кетирилген катанын наркы кампаниялар үчүн өтө чоң – сапатсыз маалыматтарды чогултуу алардын бизнесине зыян келтириши мүмкүн.

Ошондой эле, маалыматтарды коомдук уюмдар чогултушу жана бериши мүмкүн. Аларды жамааттар, ыктыярчылар, бейөкмөт уюмдар, эл аралык уюмдар эреже катары, өз максаттары үчүн чогултушат - мисалы, отчеттор, изилдөөлөр ж.б. үчүн чогултушат. Мисалы, бул Дүйнөлүк банк, БУУ, ДСУ жана башка эл аралык уюмдар же жергиликтүү ӨЭУлар тарабынан чогултулган маалыматтар.

Сиздердин күнүмдүк жашооңуздун натыйжасы болуп саналган маалыматтар дагы бар - смарт сааттар (кадамдар, машыгуулар ж.б.), телефонду канча убакыт колдонгонуңуз, социалдык тармактарда канча убакыт өткөргөнүңүз жана кайсы маалыматтарга көбүрөөк көңүл бурдуңуз, банк маалыматтары (канча жумшадыңыз жана канча алдыңар) жана башкалар. Долбоор ар кандай маалыматтар баарыбыз үчүн материал болушу мүмкүн экендигин көрсөтүп турат. Мындай маалыматтар баарыбыз үчүн да материал болушу мүмкүн. Алардын негизинде, ошондой эле кызыктуу долбоорлорду түзсө болот, мисалы, дата-арт форматында, алар башка адамдарды бизди курчап турган нерселерди байкоого, бир караганда көрүнбөгөн мыйзам ченемдүүлүктөрдү байкап, бул жашоодогу эң майда деталдардан да чыгармачылык ыкмаларды табууга шыктандырат.

**Маалыматтар жалпыга жеткиликтүү болбосо, эмне кылуу керек?**

Мамлекеттен суроо-талап кылуу керек. «Мамлекеттик органдардын жана жергиликтүү өз алдынча башкаруу органдарынын карамагындагы маалыматтарга жетүү жөнүндө» Кыргыз Республикасынын Мыйзамынын 10-беренесине ылайык, суроо-талаптарга жооп эки жумалык мөөнөттө берилиши керек. Маалымат алуу жөнүндө жазуу жүзүндөгү суроо-талапка жооп толук болууга тийиш, мында кызыкдар тараптын суроо-талаптын ошол эле предмети боюнча кайра кайрылуу зарылдыгы болбошу керек.

Маалыматтарды кеңири издөөнү колдонуп, мисалы google.ru/advanced\_search же datasetsearch.research.google.com/ сыяктуу онлайн кызматтарды колдонуп издөө.

Нейрондук тармактарды колдонуңуз, бирок алар дагы эле жеткилең эмес экендигин жана алардын маалыматтарын кылдат текшерүү талап кылынаарын унутпаңыз.

Скрейпинг маалыматтарды колдонуңуз. Бул үчүн кодду кантип жазууну билүүнүн деле кереги жок. Маалыматтардын скрейпинги үчүн акысыз шаймандар, мисалы Instant Data, Data Miner, Web Scraper ж.б. бар, алар программалоо көндүмдөрүн талап кылбайт.

**Кандай маалыматтарга ишенсе болот? Алардын сапаты жогору экендигин кантип билсе болот?**

1. Дайыма маалыматтарды чогултуу үчүн колдонулган методологияны окуңуз. Ал канчалык объективдүү экендигин билиңиз

2. Дайыма маалымат тармактарындагы (датасет) метамаалыматтаррды караңыз. Мына ушундай жол менен маалыматтар качан, ким, кандай булактардан жана кандай методология менен чогултулганын биле аласыз.

**Маалыматтарды алган сайын алар боюнча төмөнкү суроолорду берүү керек:**

1. Маалыматтар качан жана кайдан чогултулган?

2. Маалыматтар кайсы мезгилди камтыйт?

3. Маалыматтар кантип чогултулган?

4. Бул маалыматтарды кайсы уюм берген?

5. Уюм бул маалыматтарды кайсы булактардан алган?

6. Көрсөткүчтөр эмнени билдирет?

7. Жана ушул сыяктуу.

**Маалыматтар менен иштөөдө ар кандай мифтер бар, бирок дайыма төмөнкүлөрдү эстен чыгарбоо керек:**

Биринчиден, маалыматтар жок жерден пайда болбойт, алар реалдуу дүйнөнүн чагылуусу болуп саналат. Алардын өздөрүнүн толуктук даражасы, чектөөлөрү бар ж.б.у.с. Бул жөнүндө унутпоо маанилүү.

Дагы бир кеңири таралган жаңылыш түшүнүк - бул маалыматтар бирден-бир объективдүү булак жана аны менен талашууга болбойт. Бул таптакыр туура эмес. Реалдуу адамдар сыяктуу эле, маалыматтар 100 пайыз объективдүү болушу мүмкүн эмес.

Маалыматтар менен иштөөгө мүмкүн болушунча сын көз караш менен мамиле кылуу маанилүү. Алардан өзүң көргүң келгенди, издөөнүн кереги жок.

**Маалыматтардын сапаттуулугунун критерийлери болгондугуна карабастан, алар эч убакта идеалдуу боло албайт.**

1. The Open Definition, http://opendefinition.org/ [↑](#footnote-ref-1)