General Labeling Principles

- The goal is to identify *one* correct "semantic parse" of the given question. This implies:
  - Assuming the KB is **complete** and **correct**, then the correct semantic parse will result in correct answers.
  - Appropriate constraints (answer filters) should be added even when they don't change the answer set. For example, "*When was the last time Seahawks win the super bowl?*" → the "last time" constraint is needed even though Seahawks only won once. Another way to think about it is that whether the parse is correct for the same question but with different entities, such as "*When was the last time NE Patriots win the super bowl?*"
  - However, if a constraint does not provide additional information, then it should **not** be added. For example, "location.country.official_language" does not need the answer entity's type to be "language".
  - There may be multiple acceptable parses. Choose the best one based on your judgment (but see some common cases below).

- Implicit "now" – A lot of questions implicitly assume an "current" event. For example, "Lyndon B. Johnson" is not an acceptable answer to "*Who is the president of the US?*". Please remember to add the constraint (e.g., has no ".to" value) in the parse. Some common examples where the "now" constraint should be explicitly enforced:
  - Where does [somebody] live? (people.person.places_lived -- people.place_lived.location)
  - What team is [Chris Paul] on? (sports.pro_athlete.teams -- sports.sports_team_roster.team)
  - Who is the governor of Wyoming? (government.governmental_jurisdiction.governing_officials)
  - Where do the Blackhawks play? (sports.sports_team.venue -- sports.team_venue_relationship.venue)
    Where do the Fla Marlins play? (sports.sports_team.venue -- sports.team_venue_relationship.venue)
    [Do not choose "sports.sports_team.arena_stadium" because it's depreciated.]

- Similarly, if a question uses the past tense or present/past perfect, then it's preferred to choose all the historical cases/events. For example,
  - What team did [Brandon Roy] go to?
  - What team was Chris Paul on?

- Sometimes, a question will both say "now" and a year. In this case, ignore the "now" and label the question for the year. For example,
  - "who does allen iverson play for now 2010": Label it as if it were "who does allen iverson play for 2010".

- Information that can be derived trivially from the answer entity can be used. This is mostly related to the time and simple type. For example, "1984 NBA draft" is a fine answer to "*When was Michael Jordan drafted?*" (sports.drafted_athlete.drafted – sports.sports_league_draft_pick.draft). Similarly, a date is OK to be used as an answer to "how old" questions.

- Some predicates are not "official" Freebase predicates. Try to avoid them whenever possible.
  - [user.***.***] user.szaijan.fantasy_football.player.position → "*What position does Terrell Owens play?*"
  - [base.schemastaging.***] base.schemastaging.athlete_extra.salary –base.schemastaging.athlete_salary.team → "*Which team did Chris Paul play for?*"

- Sometimes the question has additional modifiers like "exactly" or "other" in it. Please just ignore them whenever appropriate. For example,
  - "What other books did Jules Verne wrote?"
  - "Where is Spain exactly located?"
  - "What region is Thailand located in?"

- Sometimes the relations in KB do not match the question exactly, but do *entail* the question. Try to choose the most matched relations. If the semantic difference between the chosen relation and question is too big, remember to put some comments (more on comments later).

- We may need some "pseudo filters" as the type of the answer entity can be easily derived from the name. For now, use "missing filter" or "missing filter value" button to report this.

- o "Which high school did [somebody] go to?" → The entity usually has "high school" in its name.

- When the correct parse is not representable using the current framework, please take a note and skip it. A common case is the "union" (more on comments later).
    - o The question is a combination of two questions. For example, "*Who is Franklin and what did he invent?*"
    - o The complete answers will need more than one chain, such as "*Where does the [UK] trade with?*" Technically, we need both "places_imported_from" and "places_exported_to".
    - o Another possible example is "*What did <somebody> play?*" If the person is regular actors for both movies and TV shows (e.g., George Clooney), then we may need both "film.actor.film - film.performance.film" and "tv.tv_actor.starring_roles - tv.regular_tv_appearance.series".

Common Types of Questions

- *Language spoken*
    For questions like "*What language do people in Spain speak?*" Generally, both "location.country.languages_spoken" and "location.country.official_language" are considered as correct inf. chains. However, if there are certain "quantifiers" in the question, then sometimes only one of them is correct. For example,
    - o "What *main* language do people in Spain speak?" → location.country.official_language
    - o "What is the *most common* language in Spain?" → location.country.official_language
    - o "What do *most* Nigerians speak?" → location.country.official_language
    - o "What are *all* languages spoken in Spain?" → location.country. languages_spoken

- Where is [Portuguese] language spoken?
    Both inference chains are allowed:
    - o language.human_language.countries_spoken_in
    - o language.human_language.main_country

- What is [someone] *famous* for?
    This type of question is tricky. Below are some of the acceptable predicates you could use to answer the question. Choose the best relation (from this list or others) based on your own judgement, but remember to leave a comment (i.e., "D" for description).
    - o people.person.profession (e.g., "what is kobe bryant famous for")
    - o common.topic.notable_for (e.g., "what was irving langmuir famous for")
    - o government.politician.government_positions_held – basic_title (e.g., "who is bill clinton famous for")
    - o base.activism.activist.area_of_activism (e.g., "what is dr jack kevorkian famous for")

- What did [someone] *do/accomplish*?
    This type of question has similar issues. Typically, the profession is a fine answer, but other predicates can be allowed. Choose the best one based on your own judgement, but remember to leave a note ("D" for description).
    - o people.person.profession (e.g., "what did ernest hemingway do")
    - o base.activism.activist.area_of_activism (e.g., "what did mary harris jones accomplish")
    - o sports.pro_athlete.teams – position (e.g., "what does cam newton do" → "Quarterback")

- Geographical location granularity
    You can be flexible on the *granularity* of the location. For example, the answer to the question "*Where are Facebook headquarters located?*" can be either
    - o organization.organization.headquarters - location.mailing_address.citytown
    - o organization.organization.headquarters - location.mailing_address.state_province_region
    Because you will only choose one parse, the more precise one (e.g., city) is preferred.

- *Marriage* (people.person.spouse_s – people.marriage.spouse)
    The *spouse* relation defined in Freebase is a bit tricky, as it could be either spouse or domestic partner. To correctly answer this type of question, the "type of union" constraint on the CVT node **needs** to be specified. Examples:
    - o Who has Brad Pitt been married to? (people.marriage.type_of_union = "Marriage")

- Who played <character> on <show/movie>?
  Because the character usually occur in only one show, the show constraint can be easily missed. Please remember to include the <show/movie> in the constraint. Example:
  - *Who has played Edward Quartermaine on General Hospital?*
    Inf Chain: tv.tv_character.appeared_in_tv_program -- tv.regular_tv_appearance.actor
    Constrint: tv.regular_tv_appearance.series = "General Hospital"

- Who did [Scarlett Johansson] date?
    For the celebrity dating question, either of the following are allowed:
  - base.popstra.celebrity.dated -- base.popstra.dated.participant
  - celebrities.celebrity.sexual_relationships -- celebrities.romantic_relationship.celebrity

- What drugs does [charlie sheen] do?
    Acceptable inf. chains:
  - celebrities.celebrity.substance_abuse_problems -- celebrities.substance_abuse_problem.substance
  - base.popstra.celebrity.substance_abuse -- base.popstra.substance_abuse.substance
    Rejected inf. chain:
  - base.popstra.celebrity.rehab_stay -- base.popstra.rehab_stay.substance_abused

- What religion does [tom cruise] follow?
    Acceptable inf. chains:
  - people.person.religion
  - base.popstra.celebrity.religion -- base.popstra.religion_choice.religion

- Where is [Okemos Michigan]?
  For "where is [city]" question, either of the following are allowed without constraints.
  - location.location.containedby
  - location.hud_county_place.county
    However, in general, the intent of this kind of question is not very clear. The GPS location seems to be a fine answer as well.

- There are several questions related to airports. They generally fall into these two types:
  - Which airport to fly into [Chicago]?
    - [OK] travel.travel_destination.how_to_get_here - travel.transportation.transport_terminus
      & Type = Airport
    - [OK with "?" comment] location.location.nearby_airports (military airports are included sometimes)
  - What airport in Paris?
    - [OK] location.location.nearby_airports

- Military (participate vs. lead)
  "What wars did Robert E Lee served in?", "What battles did Robert E Lee participate in?"
  - [OK] military.military_person.participated_in_conflicts
  - [OK with ?] military.military_commander.military_commands -- military.military_command.military_conflict
    "What wars did Robert E Lee lead?"
  - [NO] military.military_person.participated_in_conflicts
  - [OK] military.military_commander.military_commands -- military.military_command.military_conflict

- "Where was [Mary Todd Lincoln] from?" or "What country is Nicki Minaj from?"
  Either of these is okay, though neither fits perfectly:
  - [OK] "people.person.place_of_birth"

- o [OK with ?] "people.person.nationality" (even though it could be more than one countries)

- Championship
  The predicate "sports.sports_team.championships" covers different kinds of championships of professional sports. For instance, when applied to football, the relation can mean the Super bowl or NFC/AFC championship games. Note that the relation tells who won the championship, not who played in it. Questions like this include:
  - o When did the raiders win the superbowl?
  - o When did the Yankees win the world series?
  - o When did Blackhawks win the Stanley cup?

- Songs -- "What are all the songs Nicki Minaj is in?"
  - o [OK] music.artist.track_contributions -- music.track_contribution.track
  - o [OK] music.artist.track

Common Questions with Unclear Parses

- *Where did [somebody] grow up? Where was [somebody] raised?*
  - o [OK with ?] "people.person.place_of_birth"
- *What sports do they have in [country]?*
  - o There isn't the right relation in Freebase, but can be kind of inferred by "sports.sports_team_location.teams" It's OK to choose this relation and put ?? as comment.
- *What are the major cities in France?*
  - o Treat it as *"cities in France"* and select "missing filter"
- *Where did Vietnam originate?*
  - o Treat it as "missing relation"
- *What happened to [Justin Bieber] in 2012?*
  - o Choose an event, but leave a ? in the comment
- *What is [entity]?*
  - o Note this as a description question.
- *How many xxx?*
  - o It's ok to list all the answers instead of doing a "count of" aggregation.

Comment Field

Submit successfully?	Comment	Description	Example
Yes	'?' or "? " with comment	(Almost accept) There is a *minor* mis-match between the parse and question.	"What airport is in Seattle?" → loca-tion.location.nearby_airports
Yes	'??' or "?? " with comment	(Almost reject) There is a *major* mis-match between the parse and question.	"What sports do they have in France?" → sports.sports_team_location.teams
Yes	'D' or "D " with comment	It's really needs a paragraph to fully an-swer this question.	Famous for, best known for, do/accom-plish, etc.
No	'D' in "relation missing" field (later auto-added "W" to comment in this case)	It can be answered by "com-mon.topic.description".	"What is Magna Carta?" "Who is Elton John?"
No	'F' or "F " with comment	"Other Failure" – the current represen-tation language cannot represent it.	"What did JK Polk do before he was pres-ident?"
No	"F UNION"	A specific type of "F" that it needs to be a union of two answer sets	"Where does UK trade with?" → Needs both "imported_from" and "exported_to"

No	"F MQ"	A specific type of "F" that the question is in fact multiple questions	"Who is Franklin and what did he invent?"
No	'E' or "E " with comment	The labeling tool encountered an error.	System error in processing.
No	'B' or "B " with comment	Bad question. Either the meaning of the question cannot be interpreted or the KB is not good to answer the question.	"what vegetables can i plant in november in southern California" "what would ap xin zhao do"
No	'N' or "N " with comment	The topic entity is not in Freebase	"who owns mazda canada" (there is no *Mazda Canada, Inc* in Freebase, and Mazda does not contain information about subsidiaries)
No	'M' or "M " with comment	The answer is derivable from items in Freebase, but not an actual entry in Freebase. For example, you need to subtract two dates to answer a question about a time span.	"how old was woodrow wilson when he was elected president" (we can get the election and the birth and subtract them, but not answer just from an entry in Freebase).