

Oblig 2

Ketil Kvifte(kkv006)

11. april 2016

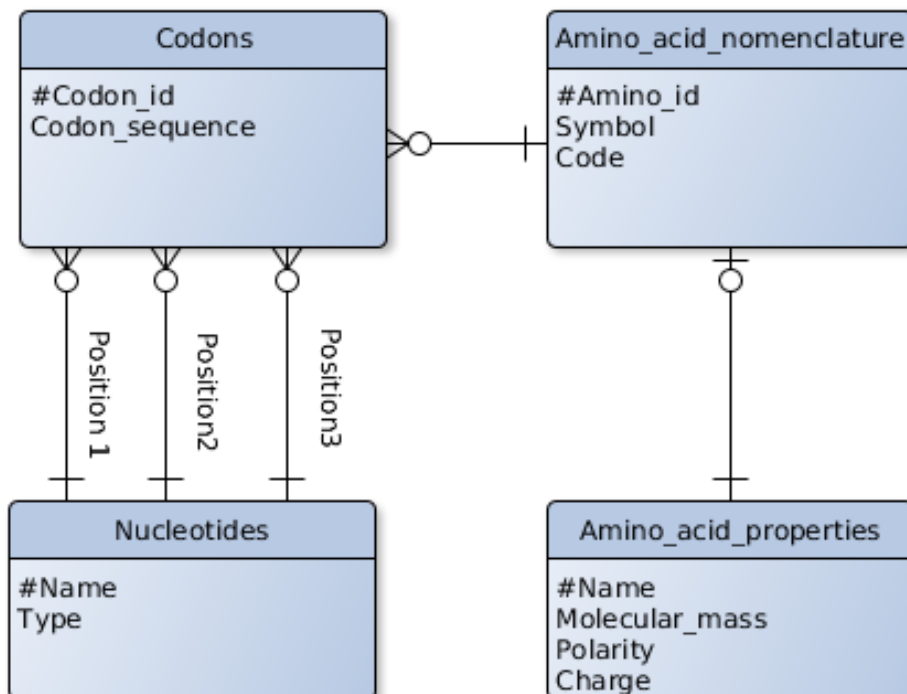
Generell kommentar til oppgåva: Det er mogleg å definere 2NF og 3NF ved bruk av kandidatnøkler, eller primærnøkler. Eg gjorde alt med kandidatnøkkel-definisjon først, men skreiv om oppgåva til å bruke primærnøkkeldefinisjon sidan det er det forelesingsnotatene og læreboka gjer. Diverre er ikkje læreboka eller Forelesingsnotata presise på definisjonen.

Læreboka/Forelesingsnotata definerer ikkje om $T(\#A, B, C, D, E), (B,C) \rightarrow (D,E)$ er brot på 3NF. Dvs at eg svarer på oppgåva som om det er eit brot. Ein har same problem med $T(\#(A,B), C, D), (C,D) \rightarrow (A,B)$ tilfredsstiller BCNF, men ikkje 2NF og dermed heller ikkje 3NF i primærnøkkeldefinisjonen. Dvs når både (A,B) og (C,D) er kandidatnøkler.

Forelesingsnotata definerer at BCNF må vere 3NF også. Difor vil eg i primærnøkkeldefinisjon eg bruker også sjekke 2NF og 3NF.

1

Eg vel å representere $\text{Codons} \rightarrow \text{Nucleotides}$ med 3 “mange-til-ein”-forhold i staden for “mange-til-mange” fordi det er meir presist.



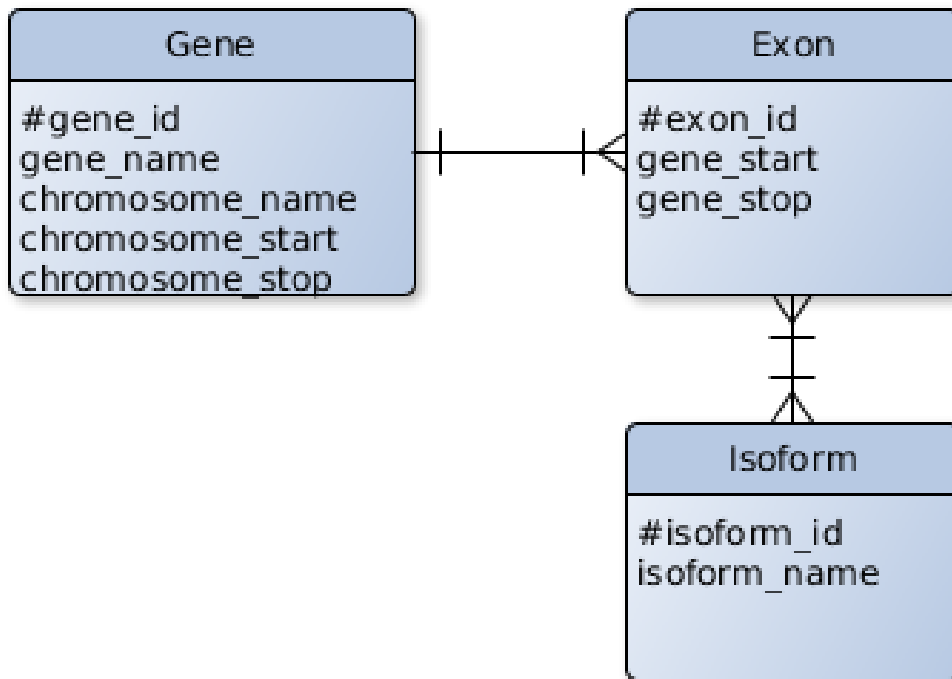
2

i)

Gene, Exon og Isoform er entiteter. Chromosome kan vere entitet, eller attributt i Genes.

ii)

Eg vel å la Chromosome vere attributt i Genes.



iii)

Gene(#gene_id*, chromosome_name)
Gene_name(#gene_id*, gene_name)
Gene_coordinate(#gene_id*, chromosome_start, chromosome_stop)
Exon(#exon_id*, gene_id*)
Exon_coordinate(#exon_id*, gene_start, gene_stop)
Isoform(#isoform_id, isoform_name)
Isoform_Exon(#(isoform_id*, exon_id*))

Forklaring Mange av tabellene har kun 2 verdier. Dvs at kriteriene for 2NF og 3NF begge er tilfredsstilt. Transitiv avhengighet gjelder ikke der fordi det er $A \rightarrow B \rightarrow C$ der $C=A$. For alle bortsett fra isoform_exon gjelder ikke partiell avhengighet fordi det ikke kan finnes ekte delmengder til primærnøkler på eit element, som samtidig er ikke-tom. For Isoform_Exon så vil ikke isoform_id implisere exon_id fordi ein isoform kan bestå av flere exon. På same måte vil ikke exon_id implisere isoform_id fordi eit exon kan inngå i flere isoformer.

I Gene_coordinate så er primærnøkkelen ikke samansett som betyr automatisk 2NF. I tillegg vil verken chromosome_start, chromosome_stop eller (chromosome_start, chromosome_stop) implisere nokon av dei andre kolonnene, og så den er heller ikke transitiv.

i)

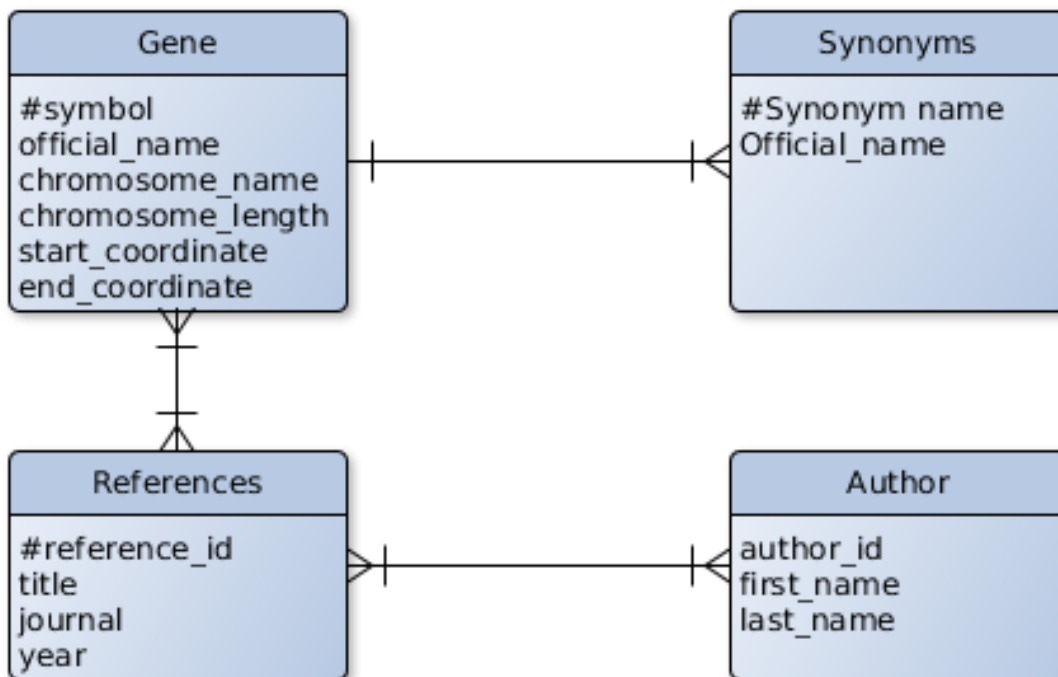
I skildringa så blir alt skildra som entiteten genome

Det gjev også meining ut frå skildringa med Gene, Chromosome, Gene_synonyms, References, Author.

Eit tredje alternativ er Gene, Gene_synonyms, References, Author.

ii)

Eg vel å lage diagram av det tredje alternativet, fordi det ser ut som gjev meir meining for oppg iii). Antar at alle gen har minst ein publikasjon og alle publikasjoner i databasen omtaler minst eit gen. Antar at alle i author har vore med på minst ein publikasjon. Antar at synonym namn er unike.



iii)

Gene(#(chromosome_name, start_coordinate), symbol, official_name, chromosome_length, end_coordinate, gene_reference_id*)

Synonyms(#Synonym_name, Official_name)

References(#reference_id, title, journal, year, gene_reference_id*, author_group_id*)

Author_group(#reference_id*, author_id*)

Author(#author_id, first_name, last_name)

Desse er automatisk 1NF fordi det kun er tillatt med atomære verdier. Det er ikkje 2NF fordi i Gene så er (chromosome_name, start_coordinate) primærnøkkel, og chromosome_name aleine bestemmer chromosome_length.

Alternativ til unike synonym_name er kombinasjonen (Synonym_name, Official_name) som primary key i synonyms.

iv)

Gene(#symbol*, chromosome_name, gene_reference_id*)
Gene_name(#symbol, official_name)
Gene_start(#symbol*, start_coordinate)
Gene_stopp(#symbol*, stop_coordinate)
Chromosome(#chromosome_name, chromosome_length)
Synonyms(#Synonym_name, Official_name)
References(#reference_id, title, journal, year)
Author_group(#reference_id*, author_id*)
Author(#author_id, first_name, last_name)

Forklaring Ingen av dei har samansett primærnøkkel, dvs minst 2NF.

I Gene er det kun primærnøkkel som impliserer andre verdier, dvs 3NF.

I Gene_name, Gene_start, Gene_stop, Chromosome, Synonym, Author_group er det kun to verdier, dvs at for alle $A \rightarrow B \rightarrow C$, så gjeld $A=C$. Dvs minst 3NF

Dersom ein journal ikkje aksepterer fleire artiklar, med same tittel i same år, så vil (title, journal, year) kunne implisere reference_id, men siden det er primærnøkkel gjer det ingenting. Dersom dei aksepterer fleire artikkler er det heller ikkje eit problem.

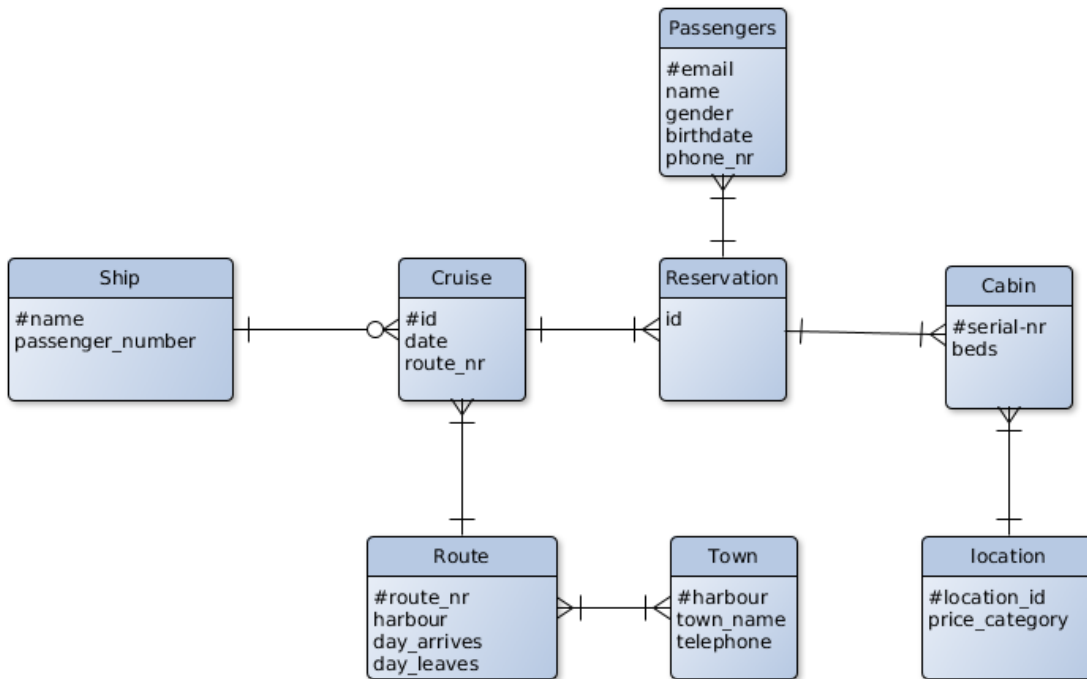
Alle med 2 maks kolonner er opplagt BCNF, fordi alle kolonner som impliserer den andre er kandidatnøkkel, og begge kolonner utgjer til saman ein supernøkkel.

Gene er BCNF fordi verken chromosome_name, gene_reference_id eller kombinasjonen, (chromosome_name, gene_reference_id) impliserer andre kolonner. Alle set som impliserer andre kolonner enn seg sjølv inneheld primærnøkkelen, som er ein kandidatnøkkel.

Dersom (title, journal, year) impliserer reference_id, så er (title, journal, year) ein kandidatnøkkel, altså supernøkkel. Dersom (title, journal, year) ikkje impliserer reference_id så er det kun primærnøkkelen som er kandidatnøkkel. Dvs at det kun er superset som inneheld den som impliserer andre. I begge tilfeller så vil alle set av kolonner som impliserer andre kolonner enn subset, vere supernøkler og difor BCNF.

I author så vil verken first_name eller last_name andre kolonner, heller ikkje kombinert. Dvs at alle set som impliserer andre kolonner enn seg sjølv inkluderer primærnøkkelen som er einaste kandidatnøkkel, og difor vil alle dei setta vere supernøkler.

Cruiset started ved dag 0, altså vil cruise starte ved byen med $\text{day_leaves} = 0$. At cruise-skipa har 4-8 dekk trengst ikkje i E/R diagrammet, det kjem først inn under constraints for DBHS. Ein finn ut når skipet reiser frå ein by ved å ta $\text{date} + \text{day_leaves}$, og når den kjem ved å ta $\text{date} + \text{day_arrives}$.



i)

Same bil kan få mange oppdrag, fleire biler kan vere med i same oppdrag. Dvs at med den definisjonen må ein duplisere alle radene for kvart oppdrag, og for kvar bil.

ii)

Inkluderer her nokre opplagte funksjonelle avhengigheter $A \rightarrow A$ for å vise at dei ikkje er gløymd.

- $\text{Registration_number} \rightarrow (\text{Registration_number}, \text{Registration_year}, \text{Model}, \text{Maximum_weight})$
- $\text{Registration_year} \rightarrow (\text{Registration_year})$
- $\text{Model} \rightarrow (\text{Model})$
- $\text{Maximum_weight} \rightarrow (\text{Maximum_weight})$
- $\text{Assignment_Number} \rightarrow (\text{Assignment_Number})$
- $(\text{Model}, \text{Registration_year}) \rightarrow (\text{Model}, \text{Registration_year})$
- $(\text{Registration_number}, \text{Assignment_Number}) \rightarrow (\text{Registration_number}, \text{Registration_year}, \text{Model}, \text{Maximum_weight}, \text{Assignment_Number})$

Antar at det ikkje er funksjonell avhengighet frå $(\text{Model}, \text{Registration_year}) \rightarrow (\text{Model}, \text{Registration_year}, \text{Maximum_weight})$, men er litt usikker på om det stemmer. Dette krev kunnskap om regelverket for køyretøy for å finne ut.

iii)

Einaste kolonne som gjev $\text{Registration_number}$ er $\text{Registration_number}$. Einaste kolonne som gjev Assignment_Number er Assignment_Number . Siden desse to til saman gjev alle kolonnene betyr det at einaste minimale supernøkkel, altså kandidatnøkkel, er $(\text{Registration_number}, \text{Assignment_Number})$ fordi begge må vere med, og ingen fleire treng vere med.

iv)

Antar at type_name er unikt per type konteiner og at det er ein kandidatnøkkel. Antar at det er funksjonell avhengighet (Max_weight, Cubic_quantity) → Nightly_rate. Dvs

- Container_type(#Type_id*, Max_weight, Cubic_quantity)
- Container_name(#Type_id*, Type_name)
- Nightly_rate(#Type_id, Nightly_rate)

Alle dei over er minst 2NF fordi det ikkje er samansett primærnøkkel. Det er kun Type_id som impliserer andre kolonner i Container_type, Dvs 3NF. Dersom (Max_weight, Cubic_quantity) impliserer den siste er det ein kandidatnøkkel. Det betyr at det kun er supernøkler som impliserer andre verdier uansett om det er ein eller to minimale kandidatnøkler.

Container_name og Nightly_rate har kun to verdier, så der er det opplagt at dei oppfyller både 3NF og BCNF.

- Container(#Container_number, Type_id*)
- Customer (#Telephone_number, Address)

Container og Customer er har kun 2 kolonner så er trivielt BCNF.

For Assignment() så står det at Container_number og Assignment_number er 1:1. Det gjev ikkje meining å ha fleire biler som transporterer samme konteiner, så det er her nødvendig å endre Container_number til Containter_order, og lage ein ny tabell som matcher fleire konteinarar mot same ordre.

- Assignment(#Assignment_number, Telephone_number*, Container_order)
- Assignment_start_date(#Assignment_number*, Start_date)
- Assignment_end_date(#Assignment_number*, End_date)
- Container_order(#(Container_order*, Container_number))

Ingen av dei over har samansett primærnøkkel, dvs minst 2NF. Ingen av dei har transitiv frå primærnøkkel. Dvs minst 3NF.

Det er ingen funksjonell avhengighet frå Telephone_number, til nokon andre. Det kan også tenkes at ein Container_order kan med dei same Containerane kan bli flytta på ulike dato i forskjellige Assignments. Dvs at Assignment_number er einaste kandidatnøkkel i Assignment, og alle set som impliserer andre kolonner enn seg sjølv inkluderer denne. Dvs BCNF.

Resten har kun to kolonner og er difor automatisk 2NF, 3NF og BCNF.

- Truck(#Registration_number, Registration_year, Model, Maximum_weight)
- Assignment_truck(#(Assignment_number*, Registration_number*))

Ein må dele Truck opp i to deler for å få den over i 2NF via primærnøkkeldefinisjonen. dette fikser også 3NF. BCNF er også trivielt.

Dersom antagelsen min om funksjonell avhengighet i ii) ikkje stemmer så må ein dele den opp slik:

- Truck (#Registration_number, (Registration_year, Model)*)
- Assignment_truck(#(Assignment_number*, Registration_number*))
- Truck (#(Registration_year, Model)*, Maximum_weight)