**Types of Questions and Key points**

Goal is to answer the question and go to the key frame according to the description.  
The person may have seen the video once and might want to scroll to the designated point instead of searching for it himself.

Note: Initially, the video starts with some slightly different scenes, not relevant to cooking (like mobile and laptop in hand)

We have to prepare a list of kitchen items commonly used / available / visible, and common actions performed. We can see other datasets too.

1. Pasta Salad  
   <https://www.dropbox.com/s/ry0yisrmrwmzgyx/OP01-R01-PastaSalad.mp4?dl=0>  
     
   Questions:  
   A) What did he do before putting pasta on stove?  
   Ans. Pour water in the vessel by tap water.  
     
   B) What all vegetables does he take down out of fridge?  
   Ans. First tomato, then X vegetable, then Y vegetable.  
     
   C) How many tomatoes is he cutting?  
   Ans. Half  
     
   D) How hot was the water when capsicum was cut?  
   Ans. Mildly hot.  
     
   E) When the water started boiling, what vegetable is being cut?  
   Ans.  
     
   F) When he was taking out item X out of fridge, what else did he do?  
   Ans. Searching / arranging  
     
   G) When he was taking X out of fridge, what else did he take out of fridge along with that?  
   Ans. Something / Nothing.  
     
   H) After / Before putting all the vegetables, how many minutes is he boiling the whole contents?  
   Ans. X minutes.  
   Ans.2. He didn’t at all put vegetables in the vessel. He mixed vegetables in the end.  
     
   I) What did he do with vegetables after cutting them?  
   Ans. He mixed them with pasta in the end.  
     
   Key frames:  
   A) Frames where recipe or steps are written.  
   B) Frames where there is something on the wall.  
   C) Frames where fridge is opened.  
   D) Cutting off the tip of tomato. (I want to see how he cuts it off)  
   E) When do we see the water boiling?  
   G) When did he turn the stove off?  
   H) When did he put oil on pan? (There is no such frame)
2. Turkey Sandwich  
   <https://www.dropbox.com/s/zro9p7hoz9z3r81/OP01-R02-TurkeySandwich.mp4?dl=0>  
   Some questions and frame detection similar to above.  
     
   Questions:  
   A) What vessel is he cooking the turkey in / using?  
   Ans. X  
     
   B) What all things / contents is he putting inside the sandwich?  
     
   C) How much time is the stove on?  
   Ans. No stove on.  
     
   D) After taking beer out of fridge, what is he cutting?  
   Ans. There is no beer in this video.  
     
   Key Frames:  
     
   A) Opening the mayo  
   B) Opening sauce  
   C) When putting the sauce
3. Bacon and Eggs  
   <https://www.dropbox.com/s/gxt3hf0mr5nd3u5/OP01-R03-BaconAndEggs.mp4?dl=0>  
   Some questions similar to above two.  
     
   Questions:  
   A) Does he pour milk and put eggs in them, or the other way round?  
   Ans. Other way round. Break eggs then pour milk.  
     
   B) How many eggs broken in total?  
   Ans. 2  
     
   C) Stirring after pouring milk, or before?  
   Ans. After most of the time.  
     
   Key Frames:  
     
   A) Putting the oil on the pan.  
   B) Breaking the egg.  
   C) Pouring milk.  
   D) Opening the box of cream cheese.  
   C) Spreading butter to bagel for first time. (If there are multiple instances, first time should be shown.)
4. Continental Breakfast  
   Here, multiple dishes are prepared.  
   <https://www.dropbox.com/s/wduykqzij6ghdmp/OP01-R04-ContinentalBreakfast.mp4?dl=0>  
     
   Questions:  
     
   A) What vessel is he spreading the butter on and putting the honey?  
   Ans. Bowl.  
     
   B) What did he just do before applying butter on the bread?  
   Ans. Kept water on the stove.  
     
   C) What all overall items did he prepare? (Question maybe out of scope)  
   Ans. Tea, bread-butter and cornflakes  
     
   Key frames:  
     
   A) Opening the microwave  
   B) Settings on microwave  
   C) Reducing the intensity of the flame.  
   D) Where bread, butter and jam are there together.  
   E) Using tea bag
5. Cheese burger  
   <https://www.dropbox.com/s/teoo1ghazqs4fzv/OP01-R05-Cheeseburger.mp4?dl=0>  
     
   Questions:  
     
   A) What all vegetables were put on the bread?  
   Ans.  
     
   B) In what order is he putting things on the bread?  
   Ans. Patty, cheese, tomatoes, lettuce, sauce and mayo.  
     
   Key frames:  
   A) Show me coke bottle inside the fridge.  
   B) Show me coke bottle outside the fridge (Not there)  
     
   A) Taking out beef patty.  
   B) When one side of patty is fried.  
   C) When patty is unfried (beginning to fry)
6. Greek Salad  
   <https://www.dropbox.com/s/k549pd4yeilbl3v/OP01-R06-GreekSalad.mp4?dl=0>  
     
   Questions:  
     
   A) What all vegetables were put?  
   Ans.  
     
   B) What all dressings were put?  
   Ans.   
     
     
   Key Frames:  
     
   A) Spreading feta cheese.  
   B) Uncut vegetables  
   C) Cut vegetables
7. Pizza  
   <https://www.dropbox.com/s/anl0qicb0zbrcjl/OP01-R07-Pizza.mp4?dl=0>  
     
   Questions:  
     
   A) What is first put on the pan?  
   Ans. Mushrooms  
   B) Was microwave used or ven used?  
   Ans. Oven  
     
   Key frames:  
     
   A) Spreading the sauce on the bread  
   B) Opening the oven.  
   C) Washing hands (never in this video)  
     
     
     
   Tasks today:  
   Send Video Question Answering to Reza  
   Read key frame selection sent by Reza  
   Representation of NLP text questions / answers ()  
   Representation of Key frames  
   List out what kind of questions are answered in Video Question Answering, and what all you want to do (New things as well as old).  
   Read video question papers

**Further Tasks:**

Imp papers according to Reza (going through details of techniques):  
***Video Question Answering via Hierarchical Spatio Temporal Attention Networks*** *(Zhou Zhao, Qifan Yang)****Movieqa*** *(for ideas)*See the code and details too.3rd imp (GIFs)

See representations of MOvieQA and ***Video Question Answering via Hierarchical Spatio Temporal Attention Networks*** as well as NLP, and code  
Even Devi Parikh and Dhruv Batra representations.  
**Preparing the questions (how to generate them)**. (How movieqa have generated,standards,etc.)  
Look through data.  
Make sorted list of papers (maybe in github)  
Then later on, see key frames.  
Follow top researchers ideas.