SEMANTIC MATCHING

APPLICATION

**A&A**

Ankit Nahata (axn172930)

Arun Kumar Kanmani (axk166630)

CONTENTS

1. [PROJECT DESCRIPTION 2](#_bookmark0)
2. PROPOSED SOLUTION 2
3. IMPLEMENTATION DETAILS 3

3.0 FAQ's…….……………………………………………………………………………………………………… ……………………………………………………3

3.1 PROGRAMMING TOOLS 3

3.2 [ARCHITECTURE 4](#_bookmark1) 3.3 RESULT AND ANALYSIS……………………………………………………………………………………………………………………………..4

3.4 PROBLEMS ENCOUNTERED…………………………………………………………………………………………………………………….5

3.5 PENDING ISSUES AND IMPROVEMENT…………………………………………………………..………………………………….6

4. APPENDIX…………………………………………………...……………………………………………………………………………..…………………………….7

# **1.PROJECT DESCRIPTION**

# The aim of this project is to implement a Frequently Asked Questions (FAQs) semantic matching application that will produce improved results. This project uses a bag-of-words strategy and an improved strategy using NLP feature and techniques.

# **2.PROPOSED SOLUTION**

# The implementation of this project is carried out though four tasks.

# **Task 1**: Create a corpus of 50 FAQs and Answers.

# **Task 2**: Implement a shallow NLP pipeline and bag-of-words matching algorithm.

# **Bag-of-words creation:**

# Tokenize the FAQs and Answers into bag-of-words.

# Create a bag-of-words for each FAQ.

# Tokenize the user’s input natural language question/statement into a bag-of-words.

# **Bag-of-words matching:**

# Return the FAQ and Answer whose bag-of-words best **statistically** matches the bag-of-words from the user’s input natural language question/statement.

# **Task 3**: Implement a deeper NLP pipeline to extract semantically rich features from the FAQs and Answers.

* Tokenize the FAQs and Answers into sentences and words
* Remove stop-words
* Lemmatize the words to extract lemmas as features
* Stem the words to extract stemmed words as features
* Part-of-speech (POS) tag the words to extract POS tag features
* Perform dependency parsing or full-syntactic parsing to parse-tree based patterns as features
* Using WordNet, extract hypernymns, hyponyms, meronyms, AND holonyms as features

**Task 4**: Implement a machine-learning, statistical, or heuristic (or a combination) based approach to semantically match the user’s input question/statement to one or more FAQs.

* Run the above described deeper NLP on an user’s input natural language question/statement and extract semantic features
* Implement a machine-learning, statistical, or heuristic (or a combination) based approach to semantically match the user’s input question/statement to one or more FAQs
* Evaluate the results of at least 10 user questions/statements for the top-10 returned FAQ matches

# **3.IMPLEMENTATION DETAILS**

# **3.0. FAQ’s**

# The FAQ’s that we have selected is form the banking domain, which include some FAQ’s covering topics like account, loan, insurance, cards, investments, etc.

# All the selected FAQ’s can be found in **APPENDIX I**.

# **3.1. PROGRAMMING TOOLS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No** |  | **Library/Tool** | **Link** |
| 1 |  | Python 3.6 | <https://www.python.org/downloads/release/python-360/> |
| 2 |  | NTLK 3.2.5 | <http://www.nltk.org/> |
| 3 |  | Stanford Core NLP 3.8.0 | <https://stanfordnlp.github.io/CoreNLP/#about> |

# **3.2. ARCHITECTUREs**

# C:\Users\apd170330\Downloads\Untitled Diagram.png

# **3.3. RESULT AND ANALYSIS**

# The result or the accuracy was calculated using the **Mean Reciprocal Rank (MRR)** to analyze our search results.

The **mean reciprocal rank** is a [statistic](https://en.wikipedia.org/wiki/Statistic) measure for evaluating any process that produces a list of possible responses to a sample of queries, ordered by probability of correctness. The reciprocal rank of a query response is the [multiplicative inverse](https://en.wikipedia.org/wiki/Multiplicative_inverse) of the rank.

MRR = 1 | Q | ∑ i = 1 | Q | 1 rank i . {\displaystyle {\text{MRR}}={\frac {1}{|Q|}}\sum \_{i=1}^{|Q|}{\frac {1}{{\text{rank}}\_{i}}}.\!}

The user queries that we have selected include the following type of FAQ versions

* Same FAQ as it is.
* Jumbled words of the FAQ.
* Replacing words with wordnet features like hypernyms, hyponyms,etc.
* Includind the words from the answers only.

# **TASK 2 MRR Results:**

# Task 2 focuses on tokenizing using keywords.

# The user’s query is tokenized to words and compared to the bad of words created by FAQ’s.

# **RESULT:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sl.** | **User input** | **FAQ\_Question** | **Rank** | **Reciprocal** |
| 1 | will we get study loan tax advantage | Will we get tax benefit on the interest paid on education loan | 1 | 1 |
| 2 | difference between current and flex account | How is a Flexi Current account different from a regular current account | 1 | 1 |
| 3 | what documents should I present for buying forex card | What documents should I submit when I purchase the HDFC Bank Hajj Umrah ForexPlus card | 3 | 0.33 |
| 4 | How to find my card present | Where should I contact or write to know about the existing offers on my cards | 9 | 0.111 |
| 5 | how to put large amount of money in my card | Can I load multiple currencies on one Multicurrency Platinum ForexPlus Chip card | 0 | 0 |
| 6 | What is the maximum age to enroll for accident insurance | Is there an age limit for opting HDFC Ergo Personal Accident Insurance policy | 1 | 1 |
| 7 | what is excluded from accidental insurance protection | What do you mean by exclusions under the Accidental Protection Plan - Hospital Cash | 4 | 0.25 |
| 8 | how to purchase mutual fund | How can I Buy/Sell Mutual Funds | 2 | 0.5 |
| 9 | how much extra I pay for balance withdraw | Do I need to pay additional amount over and above Cash withdrawal Charges levied by the HDFC Bank for ATM transaction | 0 | 0 |
| 10 | What is National Electronic Funds Transfer | What is NEFT | 1 | 1 |
|  |  | TASK 2 | MRR | 0.5191 |

# **TASK 4 MRR Results:**

# Task 4 is further enhanced in Task 2, wherein we have added the features extracted in the TASK 3 and found the cos-sin similarity between the FAQ and the query. As of now we have considered the weights of all the features as same.

# **RESULT:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sl.** | **User input** | **FAQ\_Question** | **Rank** | **Reciprocal** |
| 1 | will we get study loan tax advantage | Will we get tax benefit on the interest paid on education loan | 1 | 1 |
| 2 | difference between current and flex account | How is a Flexi Current account different from a regular current account | 1 | 1 |
| 3 | what documents should I present for buying forex card | What documents should I submit when I purchase the HDFC Bank Hajj Umrah ForexPlus card | 1 | 1 |
| 4 | How to find my card present | Where should I contact or write to know about the existing offers on my cards | 2 | 0.5 |
| 5 | how to put large amount of money in my card | Can I load multiple currencies on one Multicurrency Platinum ForexPlus Chip card | 8 | 0.125 |
| 6 | What is the maximum age to enroll for accident insurance | Is there an age limit for opting HDFC Ergo Personal Accident Insurance policy | 1 | 1 |
| 7 | what is excluded from accidental insurance protection | What do you mean by exclusions under the Accidental Protection Plan - Hospital Cash | 4 | 0.25 |
| 8 | how to purchase mutual fund | How can I Buy/Sell Mutual Funds | 2 | 0.5 |
| 9 | how much extra I pay for balance withdraw | Do I need to pay additional amount over and above Cash withdrawal Charges levied by the HDFC Bank for ATM transaction | 1 | 1 |
| 10 | What is National Electronic Funds Transfer | What is NEFT | 1 | 1 |
|  |  | TASK 4 | MRR | 0.7375 |

# **3.4. PROBLEMS ENCOUNTERED**

# The following problems were encountered:

# 

# While extracting the POS tags for tokens, we indexed it with just tags which was not efficient since the words may be semantically different but the tags were same.

# While extracting the Hypernyms we may get the most popular synset of the first tag which may be NOUN, VERB, ADJECTIVE or ADVERB. But, we need the most popular synset of the most appropriate tag. So by passing the tag of the particular word to the synset function which extracts the most popular synset we got it resolved.

# One of the potential improvements tried is to perform word sense disambiguation for head words. However, for some of the sentences the Stanford Core NLP Dependency Parser returns a null set for the same.

# As we had only 50 FAQ’s therefore it was hard for us to implement the Machine Learning model so we went with the statistical method of cos-sin similarity. If we had more questions we could have used Decision Tree or SVM.

# 

# **3.5. PENDING ISSUES AND IMPROVEMENTS**

# Some of the queries showed no improvement in Task 4. We need to rectify this issue to further enhance the results.

# We are planning to add the weights for Lemmas, Stems and Hypernyms more as compared to other features. Due to the fact that Lemmas and Stems normalize the words removing the syntactic constraints, so that we could semantically match the sentences.

# Automate the process of assigning weights based on the input query, by deciding which features are the best for the given input query sentence this could be done by a Machine Learner.

# **3.6. Screenshots of Output**

# **C:\Users\apd170330\Downloads\attachments\Capture.PNG**

# Sample output for a Query with top 10

# Bag of Words

# C:\Users\apd170330\Downloads\attachments\Capture6.PNG

# Removed Stop Words

# C:\Users\apd170330\Downloads\attachments\Capture1.PNG

# Stemming and Lemmatizing

# C:\Users\apd170330\Downloads\attachments\Capture2.PNG

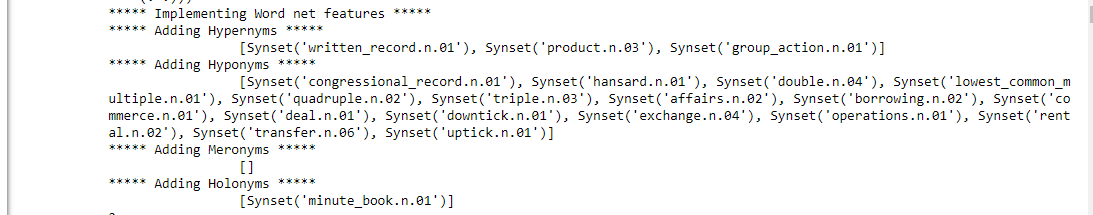
# POS tags

# C:\Users\apd170330\Downloads\attachments\Capture3.PNG

# Parse tree

# C:\Users\apd170330\Downloads\attachments\Capture4.PNG

# Wordnet Features



# **4. APPENDIX**

|  |  |
| --- | --- |
| Question | Answer |
| What details are required when I want to perform a secure IVR transaction | To perform a secure IVR transaction, you will need your 16-digit Card number, Card expiry date, CVV number, mobile number and IVR password. |
| Can the OTP be used for multiple online transactions | OTP can be used only for a single online transaction. |
| How will the online store know that I have Verified by Visa/ MasterCard SecureCode | The participating online stores will automatically recognize your Visa card number if you have registered it for Verified by Visa and a pop-up will be shown asking for your Verified by Visa password. |
| If my account is blocked, how do I unblock it | If your account gets blocked, please contact your nearest Customer Call Centre (for Credit Cards) and PhoneBanking centre (for Debit Cards). |
| I suspect someone has stolen my password or used it to make fraudulent purchases. What should I do | If you suspect any fraudulent activities, contact HDFC Bank immediately at your nearest customer Call Centre (for Credit Cards) or PhoneBanking Centre (for Debit Cards) and report the loss. |
| While trying to register I am getting error message Card already enrolled , what should I do | Your card is already registered for Verified by Visa / MasterCard SecureCode, you don't have to re-register your card. In case you don't remember your password, please click on forgot password on the Netsafe login page or on the secure page while transacting and you can reset your password. |
| What will be done with the post dated cheques if I request to change the mode of repayment/ account for my loan | Post Dated Cheques(PDCs)/Security Cheques submitted towards loan repayment that have not been encashed, will be defaced and retained by the Bank. In case, you wish to have your old Post Dated Cheques(PDCs)/ Security Cheques to be returned to you, kindly lodge a request at the Retail Loan Service Center or through PhoneBanking or write us an email at www.hdfcbank.com/services within 45 days from the date of application of change in repayment instructions. In case, we do not hear from you within the stipulated period, the defaced cheques will be destroyed. |
| How do I obtain a NOC from the bank for fitting an LPG/CNG kit on my vehicle under finance with HDFC Bank | You may obtain the NOC for fitting an LPG/CNG kit on your registered vehicle under finance with us by visiting any of our the Retail Loan Service Center and submitting the following documents: Request form for LPG/CNG NOC. Self attested copy of the Registration Certificate book, duly self attested by you, along with the Original Registration Certificate book for validation. Applicable fee of Rs. 500/- vide a cheque/payorder/demand draft favouring HDFC Bank Ltd Loan Account No. << >>. In case a bearer submits the request on your behalf, a bearer authorisation along with KYC documents (self attested copy and original for validation) of the bearer, i.e. Identity and Signature Proof like PAN Card, Aadhar Card, Driving License or Passport Copy will also need to be provided. View more |
| I don't remember my Verified by Visa / MasterCard Securecode login Id | You can use your Credit/Debit card number as your login id. The login id is required only to login to Netsafe, you don't require the login id for doing online transactions. |
| Will we get tax benefit on the interest paid on education loan | Yes, every Credila borrower or co-borrower (any one) is entitled for Tax benefit under 80E of Income Tax Act, 1961. |
| Which documents should I carry at the time of visiting HDFC bank | You need to carry an Agreement Number, Self-attested Identification Proof containing your signature to avail of any information / deliverable along with a written request and produce the documents at the Customer Service Desk. If you want to change your address / contact details, you need to submit an address change request along with a Self-attested copy of your revised address proof. The Request form for Address Change is available on the Form Centre of HDFC Bank website. View more |
| Can I give standing instructions to repay my Home Loan EMIs directly from my HDFC Bank Savings Account | Yes, you can opt for Automated Repayment and your EMIs will be directly repaid from your HDFC Bank Savings Account, thus saving you the trouble of procuring, signing and tracking post-dated cheques. |
| Who do I inform about change of name/address/status/signature, etc. | You should intimate your change of name/address/status/signature, etc. to any convenient PoS. You need to quote your PAN and submit proof (in case of a new address). You should provide at least 10 days for the change of address to take effect with all the Mutual Funds where you have investments. Please note that you should not write to the Mutual Fund or its Registrar for the change of address (unless as a designated PoS). The specified form can be obtained from the AMFI/Mutual Fund/CDSL website. All details of the holders in the Mutual Fund records will be replaced by the address details available in the CVL record. |
| What is RTGS | RTGS is the Real Time Gross Settlement which is used for the transfer of amounts of Rs. 1,00,000 or more from one bank account to another bank account (any bank located anywhere in India if attached with an Internet system) at very minimal charges. The amount is credited into the beneficiary's account in real time. For amounts less than Rs. 1,00,000 we use NEFT( National Electronic Fund Transfer). |
| What is NEFT | The National Electronic Funds Transfer (NEFT) system is a nation-wide funds transfer system to facilitate transfer of funds from any bank branch to any other bank branch. |
| How much time will it take to decision an application for Loan Against Rental Receivables | Applications would be decisioned within 7 days, and in cases which involve customer availability for completing a field investigation or title search and valuation of a property, within 25 days. |
| What is Guarantor | A Guarantor is a person who guarantees to pay for someone else's debt if he or she should default on a loan obligation. |
| What are the documents required for opening a Current Account of a sole proprietorship firm | Following documents are required to open a Current Account of a sole proprietorship entity: Proof of existence in the name of firm Proof of address in the name of firm KYC of the proprietor Any two of the below listed documents shall be obtained for establishing proof of existence. Registration certificate/license issued by Municipal authorities such as Shop & Establishment Certificate/Trade License CST/VAT/Service Tax Certificate or Letter Of Registration for CST/VAT/Service Tax Certificate/Registration document issued by Professional Tax authorities Valid Business License or Certificate Of Registration issued by State/Central Government authority (validity would include the grace period for renewal as mentioned in the certificate) RBI/SEBI Registration Certificate License issued by Food and Drug Control Authorities Import - Export certificate (IEC Code) issued by the Director General of Foreign Trade Documents for establishing proof of address Landline telephone bill/electricity bill of public & approved private operators in the state. (Bills not to be older than 4 months from the date of account opening) Property Ownership Deed i.e. title deeds of the property in the name of the firm duly stamped and registered TAN Allotment Letter issued in the name of the firm Latest property tax or water tax paid receipt/bill raised in the name of the firm Existing bank account statement or passbook of a PSU/private sector/foreign bank. Statement not older than 4 months from the date of account opening to be accepted \*\* This is only indicative list. Please visit your nearest branch for details View more |
| How is a Flexi Current account different from a regular current account | The Flexi account includes the following privileges: Flexibility of higher cash deposits (home location) and transactions anywhere (payments, collections and funds transfers), limits during peak seasons, since these are based on the Average Monthly Balance (AMB) maintained in the account. Calculation of free transaction limits based on balances maintained in the transaction month, leading to a high level of customization in tune with current business requirements. For example, if your season starts in November, you get a transaction limit based on the balance maintained by you in November, and not in October or September. Low minimum balance requirement of Rs. 75,000 per month. Competitive pricing on and faster Collection of Outstation Cheques. Up to 40 free Pay Orders and 40 free Demand Drafts per month. Free RTGS and NEFT Transactions. These privileges make a Flexi Current account different from a Regular Current account. View more |
| What transactions are included as a part of the Free Anywhere Transactions limit for a month | The following transactions comprise part of the Free Anywhere Transaction limit: Cheque deposited into the account in other cities (in local clearing) Inter-city Payable-at-Par cheque issuance. View more |
| I have an existing bank account with other Bank. Can I still open SmartUp Current Account | Yes, provided startup is meeting the criteria of date of incorporation / registration of business should not be more than five years from date of account opening. |
| Why is the principal amount and the fixed deposit maturity amount of my Fixed Deposit the same | When your Fixed Deposit is booked with the quarterly / monthly interest pay out option, the maturity amount is the same as the principal amount. The interest amount will be credited to your account at quarterly / monthly intervals. |
| Can I link accounts held in different account titles to the same Savings Account for the Sweep-In facility | No. You can only link the Sweep-In facility to the accounts with the same account title. For example, a Fixed Deposit held by A-B can be linked to a Savings Account held by A-B across branches. |
| Any minimum Balance need to be maintained in the account | No minimum balance need to be maintained as these are Zero balance accounts. |
| From where can I enquire about the status of my request to change the residential address of my HDFC Bank account | Please revert with the following details for us to look into the matter: Date of submission of request for address change The branch where the request was submitted Your Customer ID or Account number (in case, the same has not been mentioned earlier) Your new address Kindly visit the URL to access information for the complete address and contact details of our Branches as well as our PhoneBanking numbers. The primary benefit of this account is that it enables the account holder to track reimbursements you receive from your company separately. |
| What do you mean by exclusions under the Accidental Protection Plan - Hospital Cash | Exclusions are situations or conditions where HDFC ERGO is not liable to pay the benefits to the insured person in the event of an accident. |
| What is Critical Illness Insurance | Critical Illness Insurance is a policy that pays out a lump sum amount up to the sum insured upon diagnosis of a critical illness covered under the policy. |
| What do you mean by pre-existing diseases not covered under the Critical Illness - Platinum plan | A pre-existing disease is any condition, ailment, injury or related condition for which the insured person had signs or symptoms and/or was diagnosed and/or received medical advice or treatment prior to the first policy with the company. |
| Is there an age limit for opting HDFC Ergo Personal Accident Insurance policy | Personal Accident Insurance is open to everyone from the age of 18 years to 65 years. |
| What is the co pay that will need to be borne by the Life Insured in case of hospitalization in non network hospitals | The Life Insured has to bear a co-pay of 20% of the total hospital bill if they get hospitalized in any non-network hospital. |
| What are the premium paying frequencies that are available under this policy | The premium payment frequencies under this plan are Annual, Half yearly, Quarterly and Monthly. |
| What are the Policy Terms and Premium Payment Term options available under HDFC Life Uday | You have an option to select a Policy Term of 12 years or 15 years. For a 12 year term, the premium payment term is 8 years while for a 15 year term, you have the option of selecting a premium payment term of 8 years or 10 years. |
| Will HDFC ERGO General Insurance give me an advance on the basis of a damage estimate under the Motor Insurance policies | No, your vehicle must first be repaired, after which we will reimburse the amount payable to you or pay the garage directly in the case of cashless settlement under your policy. |
| What are the different types of Personal Accident Covers available for drivers and passengers | An owner driver is given Personal Accident Cover up to Rs. 2 lakhs (for cars)/ Rs. 1 lakh (for two-wheelers). This cover can also be extended to paid drivers and persons travelling in the vehicle, on a named and unnamed basis. |
| How can I Buy/Sell Mutual Funds | Once you are registered for an Investment Services Account you can start placing orders to buy/sell Mutual Funds through NetBanking services offered by HDFC Bank. Investing through Net Banking: You will have to login with your Customer ID and Internet Password to place orders for Mutual Funds on the website. You can view your Mutual Fund Holdings with their present market values on our website The cut-off time on the NetBanking Channel for the Same Day NAV for Mutual Funds transactions is 2 PM. |
| When will I be able to view the unit holdings in my account from the time of purchase | The unit holdings will be reflected within T+2 days (T being the transaction day). |
| How do I convert my existing portfolio of Mutual Fund units to an Online Mutual Fund portfolio | Currently, this facility is not available. |
| How can I update my E-mail ID in my Online Mutual Fund Account | If your E-mail ID is not updated in your Online Mutual Fund account, we request you get in touch with the nearest branch to enable us to register you for E-mail Statements of your Mutual Funds. On receipt of the form, your e-mail ID will be updated with the respective AMC within 15 working days. You will receive the holding statements on your e-mail and mailing address simultaneously after registration. |
| What do I need to submit to place an IPO/NCD order through my HDFC Securities Trading Account | Once you are registered with HDFC securities, you can successfully place an IPO/NCD/ Infrastructure Bonds order online; or Call Å¸??N trade. No filling of any form or any documentation is required if you are a registered customer. |
| What is ASBA | ASBA means 'Application Supported by Blocked Amount' as defined in clause (d) of sub Regulation (1) of regulation 2 of the SEBI (Issue of Capital and Disclosure Requirements) Regulations, 2009. ASBA is an application containing an authorization to block the application money in the Bank account for subscribing to an IPO/FPO or Rights Issue. |
| Will you give me a bill when I purchase Mudra Pure Gold Bars from HDFC Bank | We will give you an Invoice for all gold bars purchased from us. The Invoice will carry the weight, quantity and purity of the gold bars and will be signed and stamped by the Bank. |
| What is the maximum investment limit for the Section 54 EC- Capital Gain Bonds | The maximum investment limit amount is Rs. 50 Lakhs. |
| Are there transaction limits for Platinum Debit Card | Depending on the balance in your account, it is possible to withdraw up to Rs. 1 lakh at the ATM daily, and spend Rs. 2.75 lacs daily. These limits have been set for your card security. |
| Where can I use my EasyShop Titanium Royale Debit Card | Just look for the MasterCard logo when shopping. If you'd like to use an ATM, it should have the Master or Cirrus logo. And remember, you can use your EasyShop Titanium Royale Debit Card at all HDFC Bank ATMs just like a normal card. |
| Where should I contact or write to know about the existing offers on my cards | You do not have to call or write anywhere, you can know about the list of exciting offers on www.timespoints.com/debit. |
| Do I need to pay additional amount over and above Cash withdrawal Charges levied by the HDFC Bank for ATM transaction | In certain countires ATM Service providers are authorized to charge surcharge/convenience fees from the cardholders for accessing the ATM to withdraw money. These charges are levied by the ATM service provider over and above the charges levied by the HDFC Bank, and HDFC Bank bank has no control over the charges as these charges are not levied by the bank. These charges are normally displayed to the customer during the transaction and the customer has the option to either proceed with the transaction or abort if the charges are not acceptable to the customer. View more |
| Can I load multiple currencies on one Multicurrency Platinum ForexPlus Chip card | Yes, you can carry more than one currency on your Multicurrency card. Only HDFC Bank Multicurrency Platinum ForexPlus Chip card gives you the power of carrying multiple currencies on just one card. So now forget to apply for a new card every time you visit a different country. |
| What documents should I submit when I purchase the HDFC Bank Hajj Umrah ForexPlus card | Completed ForexPlus card Application Form Form A2 A photocopy of your Passport OR self attested copy of the Hajj receipt and manifest Additional address Proof in case current address of the customer is not matching with the address updated on Passport |
| What should I do if my Regalia ForexPlus Card has expired but there is an unutilised balance on the card | If your Card has expired before all funds have been utilised, you can contact HDFC Bank Phone Banking or any of our HDFC Bank branch for placing a refund request of the un-utilised balance. In case you wish you get a replacement card, you can place a request at any of our HDFC Bank branch |
| What if I forget my Prepaid NetBanking login Password | If you have forgotten the Password, then the same can be changed by selecting the option of Forgot Password available on the Prepaid NetBanking login page. |