೩

ಹಸ್ತಪ್ರತಿ ವ್ಯಾಸಂಗ

ಕಂಪ್ಯೂಟರ್ ಮತ್ತು ಕನ್ನಡ ಅಭಿವೃದ್ಧಿ

ಸಂಪಾದಕರು

ಡಾ. ವೀರೇಶ ಬಡಿಗೇರ

ಎಸ್.ಆರ್. ಚನ್ನವೀರಪ್ಪ

ಪ್ರಸಾರಾಂಗ

ಕನ್ನಡ ವಿಶ್ವವಿದ್ಯಾಲಯ, ಹಂಪಿ

HASTHAPRATI VYASANGA 3
COMPUTER MATTU KANNADA ABHIVRUDDHI
Edited by
Dr. Veeresh Badigera
S.R. Channaveerappa
Dept. of. Manuscriptology
Kannada University, Hampi
Vidyaranya 583 276
Published by
Publication Division, Prasaranga
Kannada University, Hampi
Vidyaranya 583 276
Pages: xvi+112 **Price : Rs. 60/-**
First Impression: 2002
www. kavihampi.org
©ಕನ್ನಡ ವಿಶ್ವವಿದ್ಯಾಲಯ, ಹಂಪಿ, ೨೦೦೨
ಪ್ರಕಾಶಕ
ಪ್ರಕಟಣಾ ವಿಭಾಗ, ಪ್ರಸಾರಾಂಗ
ಕನ್ನಡ ವಿಶ್ವವಿದ್ಯಾಲಯ, ಹಂಪಿ
ವಿದ್ಯಾರಣ್ಯ ೫೮೩ ೨೭೬
ಬೆಲೆ : ರೂ. ೬೦/-
ಮುಖಚಿತ್ರ : **ಕೆ.ಕೆ. ಮಕಾಳಿ**
ಅಕ್ಷರ ಸಂಯೋಜನೆ
ವಿದ್ಯಾರಣ್ಯ ಗಣಕ ಕೇಂದ್ರ
ಕನ್ನಡ ವಿಶ್ವವಿದ್ಯಾಲಯ, ಹಂಪಿ
ಮುದ್ರಣ
ಶ್ರೀವಿಶ ಮುದ್ರಣ ಆಫ್‌ಸೆಟ್ ಪ್ರಿಂಟರ್ಸ್
ಗದಗ

ಕನ್ನಡ ಪ್ರಜ್ಞೆಯನ್ನು ವಿಶ್ವಪ್ರಜ್ಞೆಯಾಗಿಸುವತ್ತ

ಸಾವಿರಾರು ವರ್ಷಗಳ ಸುದೀರ್ಘ ಮತ್ತು ಉಜ್ವಲ ಸಾಂಸ್ಕೃತಿಕ ಪರಂಪರೆಯನ್ನು ಹೊಂದಿರುವ ಕರ್ನಾಟಕದ ಆಧುನಿಕ ಇತಿಹಾಸದಲ್ಲಿ ಕರ್ನಾಟಕದ ಏಕೀಕರಣ ಒಂದು ಐತಿಹಾಸಿಕವೂ ಮತ್ತು ಮಹತ್ವಪೂರ್ಣವೂ ಆದ ವಿಶಿಷ್ಟ ಘಟನೆ. ಹಾಗೆಯೇ ಕನ್ನಡ ವಿಶ್ವವಿದ್ಯಾಲಯದ ಸ್ಥಾಪನೆ ಮತ್ತೊಂದು ಮಹತ್ವದ ಐತಿಹಾಸಿಕ ಘಟನೆ. ಕರ್ನಾಟಕದ ಇತರೆ ವಿಶ್ವವಿದ್ಯಾಲಯಗಳು ಕೇವಲ ಕೆಲವು ಜಿಲ್ಲೆಗಳಿಗೆ ಮಾತ್ರ ಸೀಮಿತಗೊಂಡಿದ್ದರೆ ಕನ್ನಡ ವಿಶ್ವವಿದ್ಯಾಲಯದ ವ್ಯಾಪ್ತಿ ಅಖಂಡ ಕರ್ನಾಟಕ ಮಾತ್ರವಲ್ಲದೆ ಕನ್ನಡಿಗ ಮತ್ತು ಕನ್ನಡ ಸಂಸ್ಕೃತಿ ನೆಲೆಸಿರುವ ಎಲ್ಲ ದೇಶ ಮತ್ತು ವಿದೇಶಗಳನ್ನೂ ಒಳಗೊಂಡಿದೆ. ಈ ಕಾರಣದಿಂದ ನಮ್ಮ ವಿಶ್ವವಿದ್ಯಾಲಯದ ದಾರಿ ಮತ್ತು ಗುರಿ ಎರಡೂ ವಿಭಿನ್ನವೂ ಮತ್ತು ವೈಶಿಷ್ಟ್ಯಪೂರ್ಣವೂ ಆಗಿವೆ.

ಕನ್ನಡ ವಿಶ್ವವಿದ್ಯಾಲಯ ಕನ್ನಡ ನಾಡು, ನುಡಿ, ಸಂಸ್ಕೃತಿ ಮತ್ತು ಜನಜೀವನದ ಸರ್ವಮುಖಗಳ ವಿಶಿಷ್ಟವಾದ ಅಂತರಂಗ ಮತ್ತು ಬಹಿರಂಗ ಸಂಪತ್ತನ್ನು ಕುರಿತು ಅಧ್ಯಯನ ಮಾಡುವ, ಸಂಶೋಧಿಸುವ ಮತ್ತು ಅದರ ಅಧ್ಯಯನದ ಫಲಿತಗಳನ್ನು ಜಗತ್ತಿನಾದ್ಯಂತ ಪ್ರಸಾರ ಮಾಡಿ ಕರ್ನಾಟಕದ ಬಗೆಗಿನ ಅರಿವನ್ನು ಜನಸಮುದಾಯದಲ್ಲಿ ವಿಸ್ತರಿಸುವ ಹಾಗೂ ಅನಂತಮುಖಿಯಾದ ವಿಶ್ವಜ್ಞಾನವನ್ನು ಕನ್ನಡಜ್ಞಾನವನ್ನಾಗಿ ಪರಿವರ್ತಿಸಿ ಅದು ಕನ್ನಡಿಗರೆಲ್ಲರಿಗೆ ದಕ್ಕುವಂತೆ ಮಾಡುವ ಮೂಲಭೂತ ಆಶಯದ ಪ್ರತಿಸಿದ್ಧಿಯಾಗಿ ಸ್ಥಾಪಿತಗೊಂಡಿದೆ. ಬೋಧನೆಗಿಂತ ಸಂಶೋಧನೆ, ಸೃಷ್ಟಿಗಿಂತ ವಿಶ್ವಂಭರ ದೃಷ್ಟಿ, ಶಿಥಿಲ ವಿವರಣೆಗಿಂತ ಅತುಳ ಸಾಧ್ಯತೆಗಳನ್ನೊಳಗೊಂಡ ಅನನ್ಯ ಅಭಿವ್ಯಕ್ತಿ, ನಾಡಿನ ಕೋಟಿ ಕೋಟಿ ಶ್ರೀಸಾಮಾನ್ಯರ ವಿವಿಧ ಪ್ರತಿಭಾಶಕ್ತಿ ಮತ್ತು ಸಾಮರ್ಥ್ಯಗಳ ಸದ್ಬಳಕೆಯ ಮೂಲಕ ಅವರ ಅಂತಃಪ್ರಜ್ಞೆಯನ್ನು ಎಚ್ಚರಿಸುವ, ವಿಕಸಿಸುವ ಕ್ರಿಯಾನ್ವಿತ ಕಾಯಕ ಇದರ ದಾರಿಯಾಗಿದೆ.

ಕನ್ನಡ ನಾಡನ್ನು ಕನ್ನಡ ವಿಶ್ವವಿದ್ಯಾಲಯದಲ್ಲಿ ನೋಡು, ಕನ್ನಡ ವಿಶ್ವವಿದ್ಯಾಲಯವನ್ನು ನೋಡದಿದ್ದರೆ ಕನ್ನಡ ನಾಡಿನ ಯಾತ್ರೆ ಸಂಪೂರ್ಣವಾಗದು, ಸಾರ್ಥಕವಾಗದು ಎಂಬಂತೆ ರೂಪುಗೊಳ್ಳುತ್ತಿರುವ ಮತ್ತು ರೂಪುಗೊಳ್ಳಬೇಕಾದ ಮಹಾ ಸಂಸ್ಥೆ ಇದು. ಕನ್ನಡಪ್ರಜ್ಞೆ ತನ್ನ ಸತ್ವ ಮತ್ತು ಸ್ವತ್ವದೊಂದಿಗೆ ವಿಶ್ವಪ್ರಜ್ಞೆಯಾಗಿ ಅರಳಿ ನಳನಳಿಸಬೇಕು; ವಿಶ್ವಪ್ರಜ್ಞೆ ಕನ್ನಡ ದೇಸೀ ಪ್ರಜ್ಞೆಯೊಳಗೆ ಪ್ರವೇಶಿಸಿ, ಪ್ರವಹಿಸಿ, ಸಮನ್ವಯಗೊಂಡು, ಸಂಲಗ್ನಗೊಂಡು, ಸಮರಸಗೊಂಡು ಸಾಕ್ಷಾತ್ಕಾರಗೊಳ್ಳಬೇಕು ಎಂಬುದೇ ಇದರ ಗುರಿ. ಈ ಗುರಿಯ ಮೂಲಕ

ಪರಿವಿಡಿ

೧. ಕಂಪ್ಯೂಟರ್ ಮತ್ತು ಕನ್ನಡಾಭಿವೃದ್ಧಿ ೧

ಪ್ರೊ. ಎಂ.ಎ.ಲಕ್ಷ್ಮೀತಾತಾಚಾರ್ಯ

೨. ಕನ್ನಡ ಭಾಷೆಯ ಕಂಪ್ಯೂಟರ್ ಸಂಸ್ಕರಣ ೬

ಮೂಲ. ಕೆ.ನಾರಾಯಣ ಮೂರ್ತಿ

ಅನು. ವಿ.ಬಿ.ತಾರಕೇಶ್ವರ

೩. ಕನ್ನಡದಲ್ಲಿ ಸಾಫ್ಟ್‌ವೇರ್ ಅಭಿವೃದ್ಧಿ ೧೬

ಬಿ.ಬಿ.ಚಿನ್ಮಯ ಕುಮಾರ್

೪. ಕಂಪ್ಯೂಟರ್ ಮೂಲಕ ಕನ್ನಡ ಹಸ್ತಪ್ರತಿಗಳ ಪುನಾರಚನೆ ಮತ್ತು ಪ್ರಸರಣ ೨೦

ಡಾ. ವೀರೇಶ ಬಡಿಗೇರ

೫. ಕನ್ನಡ ಡಾಟಾಬೇಸ್‌ಗಳಲ್ಲಿ ಮಾಹಿತಿಯ ಅಳವಡಿಕೆ ೨೫

ಬಿ.ಎ.ಶಾರದ

೬. ಕಾವ್ಯಾಭ್ಯಾಸಕ್ಕೆ ಕಂಪ್ಯೂಟರ್ ಸಹಾಯ ೩೪

ಡಾ. ಶ್ರೀನಿವಾಸ ಹಾವನೂರ

೭. ಕನ್ನಡ ಜಾನಪದ ಅಧ್ಯಯನದಲ್ಲಿ ಕಂಪ್ಯೂಟರ್ ಬಳಕೆ ೫೧

ಪ್ರೊ. ಟಿ.ಎಸ್.ಸತ್ಯನಾಥ್

೮. ಕಂಪ್ಯೂಟರ್‌ಗಳಲ್ಲಿ ಕನ್ನಡ ಕ್ರಾಂತಿ ೮೭

ಬಿ.ಬಿ.ಚಿನ್ಮಯ ಕುಮಾರ್

೯. ಕನ್ನಡದಲ್ಲಿ ಗಣಕ ಕ್ರಿಯೆ ೯೨

ಪ್ರೊ. ಆ.ಶ್ರೀಧರ

೧೦. ನಿಘಂಟುರಚನೆಯಲ್ಲಿ ಕಂಪ್ಯೂಟರ್‌ನ ಬಳಕೆ ೯೬

ಡಾ. ಎಸ್.ಎಸ್.ಅಂಗಡಿ

೧೧. ಕಂಪ್ಯೂಟರ್‌ಗಳೂ, ಕನ್ನಡದ ಅಭಿವೃದ್ಧಿಯೂ ೧೦೨

ಶಂಕರ್ ಅಜ್ಜಂಪುರ

೧೨. ಕಂಪ್ಯೂಟರ್ ಕೈಬರಹವನ್ನು ಅರ್ಥೈಸಬಲ್ಲದೇ ? ೧೦೬

ಸಿಲಿಕಾನ್ ಸಿಟಿ

೧೩. ಕನ್ನಡದಲ್ಲಿ 'ಇ-ಮೇಲ್' ೧೦೯

ಮನಸ್ವಿ

೨

ಕನ್ನಡ ಭಾಷೆಯ ಕಂಪ್ಯೂಟರ್ ಸಂಸ್ಕರಣ

ಮೂಲ : ಕೆ. ನಾರಾಯಣ ಮೂರ್ತಿ

ಅನು : ವಿ.ಬಿ.ತಾರಕೇಶ್ವರ

ಈ ಲೇಖನದಲ್ಲಿ ಭಾಷಾ ಸಂಸ್ಕರಣವನ್ನು ಕುರಿತು ಮುಖ್ಯವಾಗಿ ಕನ್ನಡ ಭಾಷೆಯ ಗಣಕ ಸಂಸ್ಕರಣವನ್ನು ಕುರಿತು ಚರ್ಚಿಸಲಾಗಿದೆ. ಭಾಷಾ ಸಂಸ್ಕರಣದ ಮುಖ್ಯವಾದ ಆಯಾಮಗಳನ್ನು ಕುರಿತು ಮೊದಲು ಸಂಕ್ಷಿಪ್ತವಾಗಿ ವಿವರಿಸಲಾಗಿದೆ. ಅನಂತರ ವಿವಿಧ ಸಂಸ್ಥೆ, ಸಂಘಟನೆಗಳಲ್ಲಿ ನಡೆದಿರುವ ಕನ್ನಡ ಭಾಷೆಯ ಸಂಸ್ಕರಣೆ ಕುರಿತ ಕೆಲಸ ಕಾರ್ಯವನ್ನು ರೇಖಿಸಲಾಗಿದೆ. ಕೊನೆ ಭಾಗದಲ್ಲಿ ಕನ್ನಡ ಭಾಷೆಯ ಸಂಸ್ಕರಣದ ಅಭಿವೃದ್ಧಿಯ ಸಾಧ್ಯತೆಗಳನ್ನು ಕುರಿತ ಕ್ರಿಯಾಯೋಜನೆಯೊಂದನ್ನು ನೀಡಲಾಗಿದೆ.

೧. ಪರಿಚಯ

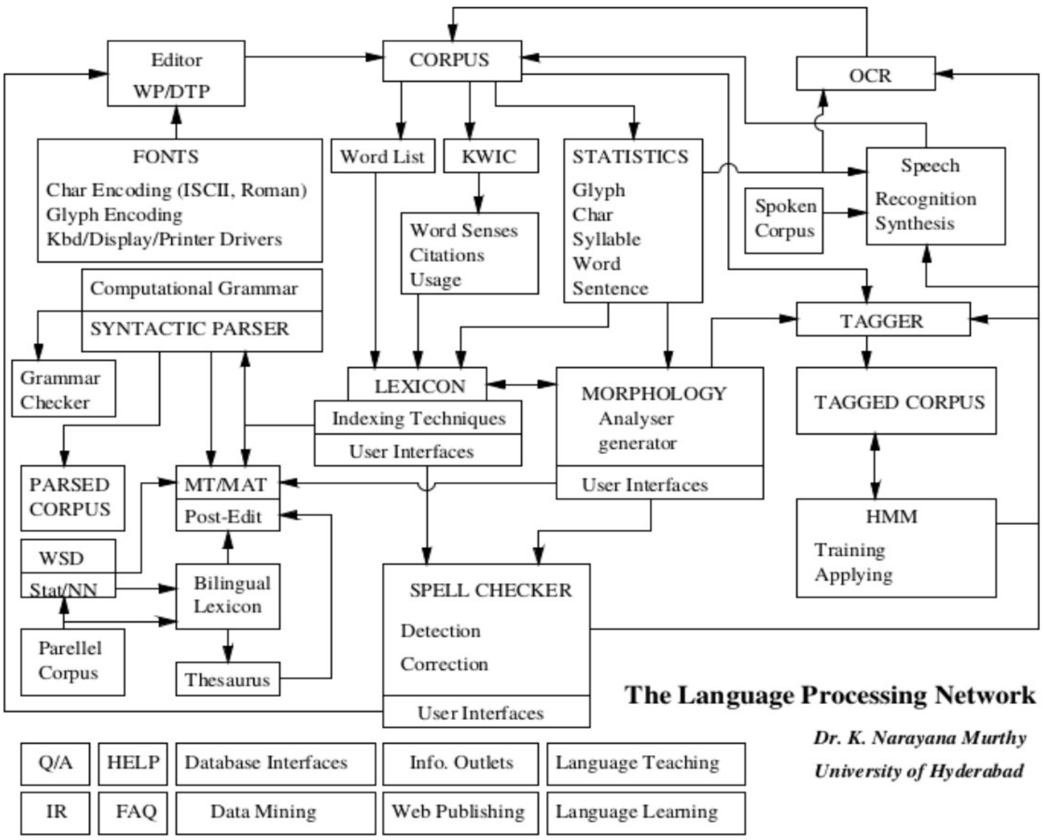
ಸಂವಹನಕ್ಕಾಗಿ ಭಾಷೆಯನ್ನು ಬಳಸುವುದು ಮನುಷ್ಯ ಜೀವಿಮಾತ್ರ. ಭಾಷೆ ನಾವು ಜೀವಿಸುವ ಪ್ರಪಂಚದ ಪ್ರತಿಬಿಂಬ. ಆದ್ದರಿಂದಾಗಿ ಅದು ವಿಶಾಲವಾದದ್ದೂ ಮತ್ತು ಬಹಳ ಸಂಕೀರ್ಣವಾದದ್ದೂ ಆಗಿದೆ. ಜನರು ತಮ್ಮ ಸಾಮಾನ್ಯ ತಿಳುವಳಿಕೆ ಮತ್ತು ಬುದ್ಧಿವಂತಿಕೆಯಿಂದ ನಿಜಜೀವನದಲ್ಲಿ ಭಾಷೆಯನ್ನು ಸಮರ್ಥವಾಗಿ ಬಳಸುತ್ತಾರೆ. ಕಂಪ್ಯೂಟರ್‌ಗಳಿಗೆ ಸಾಮಾನ್ಯ ತಿಳುವಳಿಕೆ ಇಲ್ಲ. ಕಂಪ್ಯೂಟರ್‌ಗಳನ್ನು ಬುದ್ಧಿಶಾಲಿಯಾಗಿ ಮಾಡುವ ಪ್ರಯತ್ನಗಳು ಮನುಷ್ಯನ ಬುದ್ಧಿವಂತಿಕೆಯ ಮಟ್ಟಕ್ಕೆ ಅದನ್ನು ಕರೆದೊಯ್ಯುವಲ್ಲಿ ಇನ್ನೂ ಯಶಸ್ವಿಯಾಗಿಲ್ಲ. ಆದ್ದರಿಂದಾಗಿ ಕಂಪ್ಯೂಟರ್‌ನೊಂದಿಗೆ ಸಹಜ ಭಾಷೆಯಲ್ಲಿ ಸಂವಹನ, ದೂರದ ಮಾತಾಗಿ ಉಳಿದಿದೆ.

ಆದರೆ, ಕಂಪ್ಯೂಟರ್‌ಗಿರುವ ಕೆಲವು ಅನನ್ಯ ಸಾಮರ್ಥ್ಯದಿಂದಾಗಿ ವಿವಿಧ ರೀತಿಯ ಭಾಷಾ ಸಂಸ್ಕರಣದ ಕೆಲಸವನ್ನು ನಿರ್ವಹಿಸಲು ಅದು ಸೂಕ್ತ ಸಾಧನವಾಗಿದೆ. ಸಂಬಂಧಿತ ಮಾಹಿತಿಗಳನ್ನು ಒಟ್ಟಿಗೆಯೂ ಪ್ರತ್ಯೇಕವಾಗಿಯೂ ಅದು ನೆನಪಿಟ್ಟುಕೊಂಡು, ಬೇಕಾದಾಗ ಕ್ಷಣಾರ್ಧದಲ್ಲಿ ಮತ್ತು ತಪ್ಪಿಲ್ಲದಂತೆ ಮಾಹಿತಿಯನ್ನು ಮರುನೀಡಬಲ್ಲದು. ನಮ್ಮ ಗ್ರಂಥಾಲಯಗಳಲ್ಲಿನ ಪುಸ್ತಕದ ಎಲ್ಲಾ ವಾಕ್ಯಗಳನ್ನು ನೆನಪಿಟ್ಟುಕೊಳ್ಳಬಹುದು! ಅತಿ ಉದ್ದದ ಮತ್ತು ಸಂಕೀರ್ಣವಾದ ಸರಳ ಆಜ್ಞಾಪನೆಗಳನ್ನು ಯಾಂತ್ರಿಕವಾಗಿ ನೆನೆಪಿಟ್ಟುಕೊಂಡು ತಪ್ಪಿಲ್ಲದೆ, ಅತಿ ವೇಗವಾಗಿ ನಿರ್ವಹಿಸಬಲ್ಲುದು. ಅವುಗಳಿಗೆ ಬೇಸರವಾಗಲಿ, ದಣಿವಾಗಲಿ

ತಮ್ಮ ಕೆಲಸದಲ್ಲಿಲ್ಲ. ತಪ್ಪಂತೂ ಮಾಡಲು ಸಾಧ್ಯವೇ ಇಲ್ಲ. ಗ್ರಂಥಾಲಯದಲ್ಲಿನ ಪುಸ್ತಕಗಳ ಎಲ್ಲ ಪದಗಳನ್ನು ಪಟ್ಟಿ ಮಾಡಿ, ಅವುಗಳು ಉಪಸ್ಥಿತವಾಗುವ ಸಂಖ್ಯೆಯನ್ನು ನೀಡುವುದಲ್ಲದೆ, ನಿರ್ದಿಷ್ಟ ಪದರೂಪಗಳ ಪಟ್ಟಿಯನ್ನು ಉಪಸ್ಥಿತಿಯ ಸಂಖ್ಯೆಗಳಿಗನುಗುಣವಾಗಿ ಜೋಡಿಸಬಲ್ಲದು.

ಒಬ್ಬ ವ್ಯಕ್ತಿಯ ಸಹಜ ಜ್ಞಾನವನ್ನು ಆಧರಿಸಿದ ಅಥವಾ ಸೀಮಿತ ಅಧ್ಯಯನಗಳನ್ನು ಆಧರಿಸಿ ನಡೆದ ಭಾಷೆಯನ್ನು ಕುರಿತ ತಾಂತ್ರಿಕ ಅಧ್ಯಯನಗಳು ಅಸಂಪೂರ್ಣವೂ, ಅಪರಿಪಕ್ವವೂ ಆಗಿರುವ ಸಾಧ್ಯತೆಗಳುಂಟು. ಯಾವುದೇ ರೀತಿಯ ಭಾಷಾ ಸಂಸ್ಕರಣೆಗೆ ಕಂಪ್ಯೂಟರ್‌ಗಳು ಇಂದು ಅತ್ಯಗತ್ಯವಾದ ಮೂಲಭೂತ ಸಾಧನವಾಗಿದೆ. ಮುಂದಿನ ಭಾಗದಲ್ಲಿ ಭಾಷಾ ಸಂಸ್ಕರಣ ಕಾರ್ಯದ ವಿವಿಧ ಘಟ್ಟಗಳನ್ನು ಕಾರ್ಯಗಳ ಒಂದು ಹೊರನೋಟವನ್ನು ನೀಡುವುದಲ್ಲದೆ, ಅವುಗಳ ಅಂತರ್ ಸಂಬಂಧವನ್ನು ಮತ್ತು ಆ ಕಾರ್ಯದಲ್ಲಿನ ಕಂಪ್ಯೂಟರ್ ಅಗತ್ಯವನ್ನು ಹೇಳಲಾಗಿದೆ.

೨. ಭಾಷಾ ಸಂಸ್ಕರಣದ ಕಾರ್ಯಜಾಲ



The Language Processing Network

Dr. K. Narayana Murthy
University of Hyderabad

ಮೇಲಿನ ಚಿತ್ರವು ಭಾಷಾ ಸಂಸ್ಕರಣೆಯ ಪ್ರಮುಖ ಕಾರ್ಯಗಳನ್ನು ಸಂಕ್ಷೇಪಿಸಿ ತೋರಿಸುತ್ತದೆ. ಚಿತ್ರದಲ್ಲಿನ ಬಾಣದ ಗುರುತು ಅವುಗಳ ಪ್ರಾಥಮಿಕ ಅಂತರ್‌ಸಂಬಂಧವನ್ನು ಸೂಚಿಸುತ್ತದೆ. ಕೆಳಗೆ ಮುಖ್ಯಾಂಶಗಳನ್ನು ವಿವರಿಸಲಾಗಿದೆ.

ಭಾಷಾ ಸಂಸ್ಕರಣೆಯ ಎಲ್ಲಾ ಕಾರ್ಯಗಳಿಗೂ ಆಧಾರವಾಗಿರುವುದೇ ಕಾರ್ಪಸ್ (ಪದ ಸಮುಚ್ಚಯ, ಅಂದರೆ ಸೂಕ್ಷ್ಮವಾಗಿ, ಎಚ್ಚರಿಕೆಯಿಂದ ಆಯ್ದು ಪ್ರಾತಿನಿಧಿಕ ಪಠ್ಯಗಳ ಎಲೆಕ್ಟ್ರಾನಿಕ್ ರೂಪದ ಸಂಗ್ರಹ. ಕಾರ್ಪಸ್ ಅನ್ನು ಮೂರು ವಿಧದಲ್ಲಿ ಸೃಷ್ಟಿಸಬಹುದು. ೧. ಡಿ.ಟಿ.ಪಿ ಸಾಫ್ಟ್‌ವೇರ್ ಬಳಸಿ ಅಥವಾ ಪದ ಸಂಸ್ಕರಣದ ಅಥವಾ ಸಂಕಲಕ ಬಳಸಿಕೊಂಡು ಕೀ ಬೋರ್ಡ್ ಮುಖಾಂತರ ಅಳವಡಿಸಿದ ಪಠ್ಯಗಳು, ೨. ಮುದ್ರಿತ ಅಥವಾ ಕೈ ಬರಹದ ಪಠ್ಯಗಳನ್ನು ಸ್ಕ್ಯಾನರ್ ಮುಖಾಂತರ ಸ್ಕ್ಯಾನ್ ಮಾಡಿ, ನಂತರ Optical Character Recogniation (OCR) ಸಾಫ್ಟ್‌ವೇರ್ ಮುಖಾಂತರ ಸ್ಕ್ಯಾನ್ ಮಾಡಿ ಪಡೆದ ಚಿತ್ರಗಳನ್ನು ಸಂಕಲಿಸಲು ಸಾಧ್ಯವಾದ ಪಠ್ಯಗಳಾಗಿ ರೂಪಾಂತರಿಸುವುದು. ಮತ್ತು ೩. ಮೈಕ್ರೋಫೋನ್ ಮುಖಾಂತರ ಓದಿ Speech to Text (ಮಾತಿನಿಂದ ಬರಹಕ್ಕೆ) ಸಾಫ್ಟ್‌ವೇರ್ ಮುಖಾಂತರ ಓದಿದ ಶಬ್ಧಗಳನ್ನು ಸಂಕಲಿಸಬಹುದಾದ ಪಠ್ಯವಾಗಿ ಪರಿವರ್ತಿಸುವುದು. ಈ ಎಲ್ಲಾ ವಿಧಗಳಲ್ಲೂ ಸೃಷ್ಟಿಯಾದ ಪಠ್ಯವನ್ನು ಕೈಯಿಂದ ತಪ್ಪುಗಳಿಗಾಗಿ ಪರಿಶೀಲಿಸಬೇಕಾಗುತ್ತದೆ.

ಕಾರ್ಪಸ್‌ನಿಂದ ನಾವು Txpe-token ವಿಶ್ಲೇಷಣೆಯ ಮೂಲಕ ಎಲ್ಲಾ ನಿರ್ದಿಷ್ಟ ಪದರೂಪಗಳನ್ನು ಮತ್ತು ಅವುಗಳ ಪುನರಾವರ್ತನೆಯ ಸಂಖ್ಯೆಯನ್ನು ಪಡೆಯಬಹುದು. ಈ ರೀತಿ ಪಡೆದ ಪದರೂಪಗಳನ್ನು ಅಧ್ಯಯನ ಮಾಡಿದರೆ ನಿರ್ದಿಷ್ಟ ಉದ್ದೇಶದ ನಿಘಂಟು ಅಥವಾ ಪದಕೋಶವನ್ನು ರಚಿಸಲು ದಾರಿಯಾಗುತ್ತದೆ. ಟೈಪ್ ಟೋಕ್‌ನ್ ವಿಶ್ಲೇಷಕವು ನಮಗೆ ಸಂಪೂರ್ಣ ಪದರೂಪವನ್ನು ಅಂದರೆ ಪ್ರತ್ಯಯ ಸಹಿತ ರೂಪಗಳನ್ನು ಮತ್ತು ನಿಷ್ಪತ್ತಿರೂಪಗಳನ್ನು ನೀಡುತ್ತದೆ. ಇದರಿಂದಾಗಿ ಮಾರ್ಫಾಲಾಜಿಕಲ್ ವಿಶ್ಲೇಷಣೆಯ (ಶಬ್ಧ ವಿಶ್ಲೇಷಣೆ) ಮೂಲಕ ಮೂಲಪದ ಶಬ್ಧವನ್ನು ಪಡೆಯಬೇಕಾಗುತ್ತದೆ Key-word-In-Context ಉಪಕರಣದ ಮೂಲಕ ಕಾರ್ಪಸ್‌ನಲ್ಲಿ ನಿರ್ದಿಷ್ಟ ಪದ ಬಳಕೆಯಾಗಿರುವ ಎಲ್ಲಾ ವಾಕ್ಯಗಳ ಪಟ್ಟಿಯನ್ನು ತಯಾರಿಸಬಹುದು. ಇದರಿಂದಾಗಿ ನಿರ್ದಿಷ್ಟ ಪದವು ವಿಭಿನ್ನ ಸಂದರ್ಭಗಳಲ್ಲಿ ಯಾವ ಯಾವ ಅರ್ಥದಲ್ಲಿ ಬಳಕೆಯಾಗಿದೆ ಎಂದು ಅಧ್ಯಯನ ಮಾಡಲು ಸಹಾಯವಾಗುತ್ತದೆ. ನಿಘಂಟುವಿನ ಪದಕ್ಕೆ ಅರ್ಥಗಳನ್ನು ಹೇಳುವಾಗ ಅರ್ಥಗಳನ್ನು, ವ್ಯಾಖ್ಯಾನಗಳನ್ನು ಸುಲಭವಾಗಿ ಸೇರಿಸಲು ಇದರಿಂದ ಸಾಧ್ಯವಾಗುತ್ತದೆ. ಮಾನವನ ಬುದ್ಧಿವಂತಿಕೆ ಕೆಲವು ಸಂಭವನೀಯ ಅರ್ಥಗಳನ್ನು ಮರೆಯಲು/ನೆನಪಿನಲ್ಲಿಟ್ಟುಕೊಳ್ಳದಿರಲು ಸಾಧ್ಯವಿದೆ, ಆದ್ದರಿಂದ ದೊಡ್ಡದಾದ ಕಾರ್ಪಸ್ ಆಧರಿಸಿದ ಅಧ್ಯಯನ ಬಹಳ ಜರೂರಾದದ್ದು. ಇದೇ ರೀತಿ ಕಾರ್ಪಸ್‌ನಿಂದ ಉದ್ಧರಣೆಗಳು, ಬಳಕೆ ಟಿಪ್ಪಣಿಗಳು, ಉದಾಹರಣೆಗಳು, ಉಲ್ಲೇಖಗಳನ್ನು ಸಹ ನಾವು ಪಡೆಯಬಹುದು. ಸಮರ್ಥ ಮರುಪಡೆಯುವಿಕೆಗೆ ಸೂಚಿತಂತ್ರವನ್ನು ಮತ್ತು ಬಳಸುವವರ ಅನುಕೂಲಕ್ಕೆ ಯಂತ್ರಬಳಸುವವರ ಮುಖಾಮುಖಿಯನ್ನು ಸೇರಿಸಿದರೆ ನಿರ್ದಿಷ್ಟ ಉದ್ದೇಶಕ್ಕೆ ವಿನ್ಯಾಸಗೊಳಿಸಿದ ಎಲೆಕ್ಟ್ರಾನಿಕ್ ನಿಘಂಟು ತಯಾರಾಗುತ್ತದೆ. ಎಲೆಕ್ಟ್ರಾನಿಕ್ ನಿಘಂಟುಗಳು

ಕೇವಲ ಸಾಂಪ್ರದಾಯಿಕ ನಿಘಂಟುಗಳ ಎಲೆಕ್ಟ್ರಾನಿಕ್ ರೂಪಗಳಲ್ಲ. ಪುಟಗಳನ್ನು ಅತಿ ಶೀಘ್ರದಲ್ಲೇ ಮುಗಚಬಹುದೆಂಬುದಷ್ಟೇ ಅಲ್ಲ, ಅದು ಸಾಂಪ್ರದಾಯಿಕ ನಿಘಂಟುವಿನಿಂದ ಸಾಧ್ಯವಿಲ್ಲದ ಇತರೆ ಹೆಚ್ಚಿನ ಸೌಲಭ್ಯಗಳನ್ನು ಸಹ ನೀಡುತ್ತದೆ. ಉದಾಹರಣೆಗೆ ಇದರ ಮೂಲಕ ಎಲ್ಲಾ ಕ್ರಿಯಾಪದಗಳು ಎಲ್ಲಾ ಎರಡು ಮಾತ್ರೆಯ ಪದಗಳ, ಕ್ರಿಯಾಪದ ಮತ್ತು ನಾಮಪದ ಎರಡೂ ಅರ್ಥದಲ್ಲಿ ಬಳಸುವ ಪದಗಳ, ನಿರ್ದಿಷ್ಟ ಪ್ರತ್ಯಯದಿಂದ ಕೊನೆಗೊಳ್ಳುವ ಪದಗಳ ಪ್ರತಿ ಇತ್ಯಾದಿಗಳನ್ನು ಪಡೆಯಬಹುದು. ಕಂಪ್ಯೂಟರ್‌ಗೆ ಎಲೆಕ್ಟ್ರಾನಿಕ್ ನಿಘಂಟು ಬಹಳ ಉಪಯೋಗಕ್ಕೆ ಬರುತ್ತದೆ. ಜನರ ಬಳಕೆಗೆ ಉಪಯೋಗವಾಗುವ ನಿಘಂಟುವಿಗೂ, ಕಂಪ್ಯೂಟರ್‌ಗೆ ಅಗತ್ಯವಾದ ನಿಘಂಟುವಿಗೂ ಬಹಳ ವ್ಯತ್ಯಾಸವಿದೆ. ಹೆಚ್ಚಿನ ವಿವರಗಳಿಗ ನೋಡಿ (೪).

ಪದಗಳ ಆಂತರಿಕ ರಚನೆಯನ್ನು ಅಧ್ಯಯನ ಮಾಡುವ ಮಾರ್ಫಾಲಾಜಿ ಯಾವುದೇ ಭಾಷಾ ಸಂಸ್ಕರಣೆಯಲ್ಲಿ ಬಹಳ ಅಗತ್ಯವಾದ ಭಾಗ. ಭಾರತೀಯ ಭಾಷೆಗಳು ಅದರಲ್ಲೂ ಬಹಳ ಸಂಕೀರ್ಣವಾದ ಆಂತರಿಕ ಪದರಚನಾ ವ್ಯವಸ್ಥೆ ಹೊಂದಿರುವ ಕನ್ನಡದಂಥ ದ್ರಾವಿಡ ಭಾಷೆಗಳ ಸಂಸ್ಕರಣೆಯಲ್ಲಿ ಇದರ ಮಹತ್ವ ಇನ್ನೂ ಹೆಚ್ಚಿನದು. ಮಾರ್ಫಾಲಜಿ ವಿಶ್ಲೇಷಕವು ತನಗೆ ದತ್ತ ಮಾಹಿತಿಯಿಂದ ಸಂಪೂರ್ಣ ಪದರೂಪವನ್ನು ತೆಗೆದುಕೊಂಡು, ಅದರ ಭಾಗಗಳನ್ನು ಸಂರಚನೆಯನ್ನು ವಿವರಿಸುತ್ತದೆ. ಕಾರ್ಪಸ್‌ನಿಂದ ಟೈಪ್ ಟೋಕನ್ ವಿಶ್ಲೇಷಕವು ಉತ್ಪಾದಿಸಿದ ಪದಗಳ ಪಟ್ಟಿಯನ್ನು ಮಾರ್ಫಾಲಾಜಿಕಲ್ ವಿಶ್ಲೇಷಕವು ವಿಶ್ಲೇಷಿಸಿ ಪದಗಳ ಮೂಲಪದವನ್ನು ಪಡೆಯುತ್ತದೆ. ತನ್ನ ಪರಿಷ್ಕರಣೆಗಾಗಿ ಮಾರ್ಫಾಲಜಿಕಲ್ ಭಾಗವು ಪದಕೋಶವನ್ನು ಆಧರಿಸಿರುತ್ತದೆ.

ಭಾಷಾ ಸಂಸ್ಕರಣೆಯ ವಿವಿಧ ಪ್ರಕ್ರಿಯೆಗಳಲ್ಲಿ ಅಕ್ಷರಗಳ, ಮಾತ್ರೆಗಳ, ಪದಗಳ ಮತ್ತು ವಾಕ್ಯಗಳ ಹಂತದ ಅಂಕಿ-ಸಂಖ್ಯಾ ವಿಶ್ಲೇಷಣೆ ಬಹಳ ಉಪಯುಕ್ತವಾದದ್ದು. ಇದರಿಂದ ಯಾವುದು ಮೂಲಭೂತವಾದುದು, ಸತತವಾಗಿ ಬಳಕೆಯಾಗುವುದು? ಯಾವುದು ವಿರಳವಾದುದು, ವಿಶಿಷ್ಟ ಸಂದರ್ಭಕ್ಕೆ ಸೀಮಿತವಾದುದು ಅಥವಾ ಭಾಷೆಯ ನಿಜಪ್ರತಿನಿಧಿಯಲ್ಲ ಎಂದು ಗೊತ್ತಾಗುವುದು. ಬಹಳ ಮುಖ್ಯವಾದವುಗಳ ಮೇಲೆ ಗಮನ ಕೇಂದ್ರೀಕರಿಸಲೂ, ಅಮುಖ್ಯವಾದವುಗಳನ್ನು ತೆಗೆದು ಹಾಕಲೂ ಇದು ಸಹಾಯ ಮಾಡುತ್ತದೆ. (೧೨,೧೩).

ಒಮ್ಮೆ ನಾವು ಒಳ್ಳೆಯ ಎಲೆಕ್ಟ್ರಾನಿಕ್ ನಿಘಂಟುಗಳನ್ನು, ಮಾರ್ಫಾಲಾಜಿಕಲ್ ವಿಶ್ಲೇಷಕವನ್ನು ಮತ್ತು ಜನರೇಟರನ್ನು (ಪದ ಉತ್ಪಾದಕ) ಹೊಂದಿದರೆ, ಅವುಗಳನ್ನು ಸಮಗ್ರಗೊಳಿಸಿ Spell Checker (ಕಾಗುಣಿತ ಪರಿಶೀಲಕ)ವನ್ನು ರೂಪಿಸಬಹುದು. ಕಾಗುಣಿತ ಪರಿಶೀಲಕದಲ್ಲಿ ಎರಡು ಭಾಗಗಳಿರುತ್ತವೆ : ೧. ಪತ್ತೆ ಹಚ್ಚುವ ಭಾಗ ೨. ತಿದ್ದುವ ಭಾಗ. ಪತ್ತೆ ಹಚ್ಚುವ ಭಾಗದಲ್ಲಿ ಕಾಗುಣಿತ ತಪ್ಪಾದ ಪದಗಳನ್ನು ಗುರುತಿಸಲಾಗುವುದು. ಕೊಟ್ಟ ಪಠ್ಯವನ್ನು, ಎಲೆಕ್ಟ್ರಾನಿಕ್ ನಿಘಂಟುವಿನೊಂದಿಗೆ ಹೋಲಿಸಿ ಇದನ್ನು ಮಾಡಬಹುದು.

ನಿಘಂಟುವಿನಲ್ಲಿನ ಪದಗಳನ್ನು ಸರಿಯಾದ ಪದಗಳಾಗಿ ಭಾವಿಸಿ, ಪಠ್ಯದಲ್ಲಿರುವ ಆದರೆ ನಿಘಂಟುವಿನ ಅಲ್ಲದ ಪದಗಳನ್ನು ಕಾಗುಣಿತ ತಪ್ಪಾದ ಪದಗಳು ಎಂದು ಭಾವಿಸಲಾಗುತ್ತದೆ. ಇದರಿಂದ ನಿಘಂಟುವಿನಲ್ಲಿ ಸೇರದ ಆದರೆ ಸರಿಯಾದ ಪದಗಳೂ ಕೂಡ ಕಾಗುಣಿತ ತಪ್ಪಾದ ಪದಗಳಾಗಿ ಕಾಣಿಸಲ್ಪಡುತ್ತವೆ. ಇದಕ್ಕೆ ಪರ್ಯಾಯವಾದ ವಿಧಾನವೆಂದರೆ, n-gram ಅಂಕಿ-ಸಂಖ್ಯೆಗಳನ್ನು ಆಧರಿಸಿ ಕಮ್ಮಿ ಸಾಧ್ಯತೆಗಳುಳ್ಳ ಪದಗಳನ್ನು ತಪ್ಪಾದ ಪದಗಳಾಗಿ ತೋರಿಸುವುದು. ಭಾಷಾಶಾಸ್ತ್ರ ಮತ್ತು ಸಂಖ್ಯಾಶಾಸ್ತ್ರ ಎರಡರ ತಂತ್ರಗಳನ್ನು ಬೆರೆಸಿದ ವಿಧಾನವನ್ನೂ ಅನುಸರಿಸಬಹುದು. ತಿದ್ದುವಭಾಗದಲ್ಲಿ, ಕಾಗುಣಿತ ಪರಿಶೀಲಕ ಪರ್ಯಾಯಗಳನ್ನು ಓದುಗರ ಆಯ್ಕೆಗೆ ಸೂಚಿಸುತ್ತದೆ, ನಿಘಂಟು ಆಧರಿತ ತಂತ್ರದಲ್ಲಿ ತಪ್ಪಾದ ಪದದ ಕಾಗುಣಿತಕ್ಕೆ ಸಮೀಪವಿರುವ ಇತರ ಎಲ್ಲಾ ಪದಗಳನ್ನು ಪರ್ಯಾಯವಾಗಿ ಸೂಚಿಸಲಾಗುವುದು. 'ಸಮೀಪ'ದ ಪರಿಮಾಣಾತ್ಮಕ ಮತ್ತು ನಿರ್ದಿಷ್ಟ ವ್ಯಾಖ್ಯೆಗೆ Minimum Edit Distance (ಕನಿಷ್ಠ ಸಂಕಲನ ದೂರ) ಮುಂತಾದವುಗಳನ್ನು ಬಳಸಲಾಗುವುದು. ಇದಲ್ಲದೆ ಬಳಸುವವರು ಸೂಚಿತ ಪರ್ಯಾಯಗಳನ್ನು ತಿರಸ್ಕರಿಸಿ, ಪದವನ್ನು ಇದ್ದ ಹಾಗೇ ಒಪ್ಪಿಕೊಳ್ಳಲು ಹೇಳಬಹುದು ಮತ್ತು ತಮ್ಮ ವ್ಯಕ್ತಿಗತ ನಿಘಂಟುವಿಗೆ ಸೇರಿಸಲೂ ಆದೇಶ ನೀಡಬಹುದು. ಯಾವುದೇ ಪದ ಸಂಸ್ಕರಣ ತಂತ್ರಾಂಶಕ್ಕೆ ಕಾಗುಣಿತ ಪರಿಶೀಲಕ ಬಹಳ ಮೌಲ್ಯಯುತ ಸೇರ್ಪಡೆಯಾಗುತ್ತದೆ. ಮಾತನ್ನು ಗುರುತಿಸಲು ಮತ್ತು OCR ಗಳಲ್ಲಿ ಸಹ ಕಾಗುಣಿತ ಪರಿಶೀಲಕ ಬಹಳ ಉಪಯುಕ್ತವಾದುದು.

ಒಂದು ಪದಕೋಶ ಪದ ಬಳಕೆಯಾಗುವ ಸಂದರ್ಭವನ್ನು ಕೊಡದೆ ಅದು ಕ್ರಿಯಾಪದವಾಗೋ, ನಾಮಪದವಾಗೋ, ಗುಣವಿಶೇಷಣವಾಗಿಯೋ ಬಳಕೆಯಾಗುವ ಸಾಧ್ಯತೆಗಳನ್ನು ನೀಡುತ್ತದೆ, ಅಂದರೆ ಅದರ Parts of Speech (POS) ನೀಡುತ್ತದೆ. ಆದರೆ StatisticalPos Tagger, Hidden Mather model (ಅಡಗಿ ಕೂತ ಮಾರ್ಕೇಟ್ ಮಾದರಿ) ಅನ್ನು ಆಧರಿಸಿದ ಒಂದು ಸಾಫ್ಟ್‌ವೇರ್ (ತಂತ್ರಾಂಶ) ಅದು ಪಠ್ಯದಲ್ಲಿನ ಪದಗಳನ್ನು ಅದರ ಬಳಕೆಯ ಸಂದರ್ಭದ ವಾಕ್ಯವನ್ನು ಪರಿಗಣಿಸಿ ಅದರ ವ್ಯಾಕರಣ ಅಂಶಗಳನ್ನು ನೀಡುತ್ತದೆ. ಪಠ್ಯಗಳನ್ನು tag ಮಾಡಲು ಮಾರ್ಫಾಲಜಿ ಆಧರಿಸಿದ ಭಾಷಾ ವಿಜ್ಞಾನದ ತಂತ್ರಗಳನ್ನು ಬಳಸಬಹುದು tag ಮಾಡಿದ ಕಾರ್ಪಸ್ ಬಹಳ ಉಪಯುಕ್ತವಾಗಬಲ್ಲದು. HMM ಅನ್ನು ತರಬೇತಿಗೊಳಿಸಲು ಅದರ ಉಪಯುಕ್ತತೆ ಬಹಳವಾದದ್ದು. ಮೊದಲಾಗಿ ಒಂದು ಒಳ್ಳೆಯ Tagger, tagged ಕಾರ್ಪಸ್ ಸೃಷ್ಟಿಸುವಲ್ಲಿ ಸಹಾಯಕಾರಿಯಾಗಬಲ್ಲದು. OCR ಮತ್ತು Speech ತಂತ್ರಾಂಶಗಳಲ್ಲಿ HMMಗಳು ಬಹಳ ಬಳಕೆಯಾಗುತ್ತವೆ.

ಪಠ್ಯಕಾರ್ಪಸ್‌ಗಿಂತ ಮಾತಿನ ಕಾರ್ಪಸ್ (Speech corpus) ಮಾತನ್ನು ಸಂಸ್ಕರಿಸುವಲ್ಲಿ ಅತ್ಯಂತ ಉಪಕಾರಿಯಾದದ್ದು. ಮಾತಿನಿಂದ ಪಠ್ಯಕ್ಕೆ, ಪಠ್ಯದಿಂದ ಮಾತಿಗೆ ಪರಿವರ್ತಿಸುವ

ವ್ಯವಸ್ಥೆಗಳಲ್ಲಿ ಅಲ್ಲದೆ ಮಾತಾಡುವವರನ್ನು ಗುರುತಿಸಲು ಮತ್ತು ಭಾಷೆ, ಉಪಭಾಷೆಗಳನ್ನು ಗುರುತಿಸಲೂ ಸಹಕಾರಿಯಾದದ್ದು.

ಕಂಪ್ಯೂಟರ್ ವ್ಯಾಕರಣವು ವಾಕ್ಯಾಧಾರಿತ Parsing ವ್ಯವಸ್ಥೆ ರೂಪಿಸಲು ಅನುವು ಮಾಡಿಕೊಡುತ್ತದೆ. ಜನರ ಬಳಕೆಗೆ ರೂಪಿತವಾಗಿರುವ ಸಾಂಪ್ರದಾಯಿಕ ವ್ಯಾಕರಣಕ್ಕಿಂತ ಕಂಪ್ಯೂಟರ್ ವ್ಯಾಕರಣವೂ ಮತ್ತಷ್ಟು ವಿವರಣಾತ್ಮಕವೂ ನಿಚ್ಚಳವೂ ಆಗಿರಬೇಕಾಗುತ್ತದೆ. ಸಾಂಪ್ರದಾಯಿಕ ವ್ಯಾಕರಣವು ಬಳಕೆದಾರರಿಗೆ ಈಗಾಗಲೇ ಭಾಷೆ ಗೊತ್ತಿದೆಯೆಂಬ ಊಹೆಯ ಮೇಲೆ ನಿರ್ಮಿತವಾಗಿರುವುದರಿಂದ ವಿಶೇಷಗಳ ಮೇಲೆ ಹೆಚ್ಚು ಕೇಂದ್ರೀಕೃತವಾಗಿರುತ್ತದೆ. ಪಾರ್ಸ್, ಪಾರ್ಸ್‌ಡ್ ಕಾರ್ಪಸ್ ಸೃಷ್ಟಿಸಲು ಸಹಾಯ ಮಾಡುವುದಲ್ಲದೆ ಅದರ ಮೂಲಕ ಸಂದರ್ಭ ಆಧಾರಿತ ಕಾಗುಣಿತವನ್ನು ಚೆಕ್ ಮಾಡಲು, ಟ್ಯಾಗಿಂಗ್ ಮಾಡಲು, OCR ನಲ್ಲಿ ಮಾತು ಗುರುತಿಸಲು ಕೂಡ ಬಳಕೆಯಾಗುತ್ತದೆ.

ಭಾಷಾಂತರ ಯಂತ್ರ ವ್ಯವಸ್ಥೆ (MT) ಅಥವಾ ವಾಸ್ತವವಾಗಿ ಹೇಳಬೇಕಾದರೆ ಭಾಷಾಂತರ ಸಹಾಯಕ ಯಂತ್ರವ್ಯವಸ್ಥೆ (MAT) ರೂಪಿಸಲು ದ್ವಿಭಾಷಾ ಪದಕೋಶ, ಪಾರ್ಸರ್‌ಗಳು, ಮಾರ್ಫಾಲಜಿಕಲ್ ವಿಶ್ಲೇಷಕಗಳು ಮತ್ತು ಜನರೇಟರ್ ಇತ್ಯಾದಿಗಳನ್ನು ಬಳಸಬೇಕಾಗುತ್ತದೆ. ಸಂಪೂರ್ಣ ಯಾಂತ್ರೀಕೃತವಾದ ಉತ್ತಮ ಗುಣಮಟ್ಟದ ಭಾಷಾಂತರವು ಮುಕ್ತ ವಲಯದಲ್ಲಿ ಸಾಧ್ಯವೇ ಇಲ್ಲವೆಂಬುದು ಈಗಾಗಲೇ ಸ್ಪಷ್ಟವಾಗಿ ನಿರೂಪಿಸಲಾಗಿದೆ.

ಉತ್ತಮ ಗುಣಮಟ್ಟದ ಯಾಂತ್ರಿಕ ಭಾಷಾಂತರವು ಕೇವಲ ಅತ್ಯಂತ ಸಿಮೀತ ವಲಯದಲ್ಲಿ ಮಾತ್ರ ಯಶಸ್ವಿಯಾಗಬಲ್ಲದು. ಉದಾ: ಹವಾಮಾನ ಮುನ್ಸೂಚನೆಯ ವರದಿಗಳು. ಸಂಪಾದನ ಪೂರ್ವ ಮತ್ತು ಸಂಪಾದನೋತ್ತರ ಅಥವಾ ದ್ವಿ-ಸಂಪರ್ಕಿತ (interactive)ವಾಗಿಯಾದರೂ. ಮನುಷ್ಯ ತೊಡಗಿಕೊಳ್ಳಬೇಕಾಗುತ್ತದೆ. ಆದರೆ ವ್ಯವಸ್ಥೆಯ ಒಂದು ಕಚ್ಚಾ ಭಾಷಾಂತರವನ್ನು ನಂಬಲಸಾಧ್ಯವಾದ ಸಮಯದಲ್ಲಿ ಇತರ ಸಲಕರಣೆಗಳೊಂದಿಗೆ ಮಾಡಬಲ್ಲದು, ಇದರಿಂದಾಗಿ ಭಾಷಾಂತರ ಪ್ರಕ್ರಿಯೆಯಲ್ಲಿನ ಸಮಯ ಮತ್ತು ವೆಚ್ಚವನ್ನು ಗಮನಾರ್ಹ ಪ್ರಮಾಣದಲ್ಲಿ ತಗ್ಗಿಸಬಹುದು.

ಇದೆಲ್ಲವನ್ನು ಗಮನದಲ್ಲಿರಿಸಿಕೊಂಡು ನಾವು ಹಲವಾರು ಪ್ರಯೋಜನಕಾರಿಯಾದ ವ್ಯವಸ್ಥೆಗಳನ್ನು ರೂಪಿಸಬಹುದು, ಉದಾ: ಪ್ರಶ್ನೋತ್ತರ ವ್ಯವಸ್ಥೆ, ಮಾಹಿತಿ, ವ್ಯವಸ್ಥೆಯಲ್ಲಿನ ಮಾಹಿತಿ ಪಡೆಯಲು ಸಹಜ ಭಾಷೆಯ ಪ್ರಶ್ನೆ ವ್ಯವಸ್ಥೆ, ಪಠ್ಯಗಳನ್ನು ತೆಗೆದು ಸಾರಾಂಶ ರಚಿಸುವ ವ್ಯವಸ್ಥೆ, ಮಾಹಿತಿ ಹೊರಗೆಳೆಯುವ ವ್ಯವಸ್ಥೆ, ಭಾಷಾ ಕಲಿಕೆ/ಕಲಿಸುವಿಕೆಯ ಸಲಕರಣೆಗಳು ಇತ್ಯಾದಿ.

ಭಾಷಾ ಸಂಸ್ಕರಣೆಯೆಂದರೆ ಬಹಳ ನಿಕಟವಾಗಿ ಅಂತರ್ ಸಂಪರ್ಕಹೊಂದಿರುವ ಮಾಹಿತಿ ಜಾಲ, ಚಟುವಟಿಕೆಗಳು, ಕ್ರಿಯೆಗಳು ಮತ್ತು ಸಾಫ್ಟ್‌ವೇರು ಎಂಬುದನ್ನು ನಾವು ಗಮನಿಸಬಹುದು. ಇವುಗಳಲ್ಲಿನ ಪ್ರತಿಯೊಂದು ಅಂಶವೂ ತನ್ನ ಬೆಳವಣಿಗೆಗೆ ಮತ್ತು ಉಪಯೋಗಕ್ಕೆ ಮತ್ತೊಂದನ್ನು ಅವಲಂಬಿಸಿದೆಯೆಂಬ ಅಂಶವೂ ಮನದಟ್ಟಾಗುತ್ತದೆ. ಅದೇ

ಸಮಯದಲ್ಲಿ ಅದು ಮತ್ತೊಂದರ ಬೆಳವಣಿಗೆ ಮತ್ತು ಬಳಕೆಗೂ ತನ್ನ ಕಾಣಿಕೆ ಸಲ್ಲಿಸುತ್ತದೆ. ಪ್ರತಿಯೊಂದು ದಿಕ್ಕಿನಲ್ಲೂ ಅಭಿವೃದ್ಧಿಪಡಿಸುತ್ತಾ ಸಾಗುವುದು ಮಾತ್ರ ನಮ್ಮನ್ನು ಗುರಿ ತಲುಪಿಸಬಲ್ಲದು.

೩. ಕನ್ನಡ ಕುರಿತ ಆಗಿರುವ ಕೆಲಸಗಳ ಸಂಕ್ಷಿಪ್ತ ವಿವರ : ಕನ್ನಡ ಭಾಷೆ ಬಹಳ ಶ್ರೀಮಂತ ಮತ್ತುವೈವಿಧ್ಯಮಯವಾದದ್ದಾಗಿದೆ. ಇದುವರೆಗೆ ಸಾಹಿತ್ಯಕ್ಕಾಗಿ ಅತಿ ಹೆಚ್ಚು ಜ್ಞಾನಪೀಠ ಪ್ರಶಸ್ತಿಗಳನ್ನು ಪಡೆದಿದೆ. ಆದರೆ ವೈಜ್ಞಾನಿಕವಾದ, ತಾಂತ್ರಿಕವಾದ ಭಾಷಾ ಅಧ್ಯಯನ ಕನ್ನಡವನ್ನು ಕುರಿತು ನಡೆದಿದೆಯೋ ಎಂದರೆ ಬಹಳ ಕಮ್ಮಿ. ಕನ್ನಡದಲ್ಲಿ ಎಷ್ಟು ಪದಗಳಿವೆ. ಇವುಗಳಲ್ಲಿ ಹೆಚ್ಚು ಬಳಕೆಯ ಪದಗಳು ಯಾವುವು? ಮೊದಲನೇ ತರಗತಿ ಅಥವಾ ಐದನೇ ತರಗತಿ ವಿದ್ಯಾರ್ಥಿಯ ಭಾಷಾ ಬಳಕೆ ಏನು? ಒಂದು ಕ್ರಿಯಾಪದ ಆಧರಿಸಿ ಮಾರ್ಫಡಿಸಿ ರೂಪಿಸಲಾದ ಪದರೂಪಗಳೆಷ್ಟು? ವ್ಯಂಜನಗಳು ಗುಂಪು ಗೂಡುವಾಗ ಯಾವ ಕ್ರಮದಲ್ಲಿ ಎಷ್ಟು ಬಾರಿ ಸೇರುತ್ತವೆ? ಇನ್ನು ಏನೇನು ಕೆಲಸ ಆಗಬೇಕಾಗಿದೆ ಎಂಬುದನ್ನು ಏನು ಕೆಲಸ ಆಗಿದೆ ಎನ್ನುವುದರ ಜೊತೆ ಹೋಲಿಸಿದರೆ, ಸಧ್ಯ ನಡೆದಿರುವ ಕೆಲಸ ಅತ್ಯಂತ ಸಣ್ಣ ಪ್ರಮಾಣದ್ದು ಎಂದು ಮನವರಿಕೆಯಾಗುತ್ತದೆ. ಈ ಕೆಳಗೆ ಈಗಾಗಲೇ ವಿವಿಧ ಸಂಸ್ಥೆಗಳಲ್ಲಿ ಆಗಿರುವ ಕೆಲಸವನ್ನು ಸ್ಥೂಲವಾಗಿ ರೇಖಿಸಲು ಪ್ರಯತ್ನಿಸಿದ್ದೇನೆ. ಈ ವಿವರಣೆ ಸಮಗ್ರವಾದದ್ದು ಎಂದು ಹೇಳಲು ಸಾಧ್ಯವಿಲ್ಲ. ಇದರ ಉದ್ದೇಶ ನಾವಿಂದು ಎಲ್ಲಿದ್ದೇವೆ ಎಂದು ಸೂಚನೆ ಪಡೆಯುವುದು ಮಾತ್ರ.

ಮೈಸೂರಿನ ಭಾರತೀಯ ಭಾಷೆಗಳ ಕೇಂದ್ರಿಯ ಸಂಸ್ಥೆ (CIIL)ಮೂರು ಮಿಲಿಯನ್ ಪದಗಳ ಕಾರ್ಪಸ್ ತಯಾರಿಸಿದೆ. ಈ ಯೋಜನೆಯನ್ನು ಕೇಂದ್ರ ಸರಕಾರದ ಮಾಹಿತಿ ತಂತ್ರಜ್ಞಾನ ಸಚಿವಾಲಯ ಪ್ರಾಯೋಜಿಸಿತ್ತು. ಈ ಕೆಲಸ ಬಹಳ ಹಿಂದೆಯೇ ಆಗಿದ್ದರೂ, ಈ ಕಾರ್ಪಸ್ ಇತ್ತಿಚೆಗಷ್ಟೇ ಆಯ್ದ ಕೆಲವು ಸಂಸ್ಥೆಗಳನ್ನು ತಲುಪಿದೆ. ಈ ಕಾರ್ಪಸ್ ಅವಲಂಬಿಸಿ ಹಲವು ಸಣ್ಣ ಪರಿಕರಗಳನ್ನು ಈಗಾಗಲೇ ಅಭಿವೃದ್ಧಿ ಪಡಿಸಲಾಗಿದೆ.

ಮೇಲುಕೋಟೆಯ ಸಂಸ್ಕೃತ ಸಂಶೋಧನಾ ಅಕಾಡೆಮಿ ಸಂಸ್ಕೃತದ ಜೊತೆಗೆ ಕನ್ನಡವನ್ನೂ ಕುರಿತು ಕೆಲಸ ಮಾಡುತ್ತಿದೆ. ಪದಕೋಶಗಳು, ಪದರೂಪ ಉತ್ಪಾದಕಗಳು ಇತ್ಯಾದಿಗಳನ್ನು ರೂಪಿಸಿದೆ.

ಹೈದರಾಬಾದ್ ವಿಶ್ವವಿದ್ಯಾಲಯದ ಕಂಪ್ಯೂಟರ್ ಮತ್ತು ಮಾಹಿತಿ ವಿಜ್ಞಾನ ವಿಭಾಗವು ಕರ್ನಾಟಕ ಸರಕಾರಕ್ಕಾಗಿ ಭಾಷಾಂತರ ಸಹಾಯಕ ಯಂತ್ರ(Machine Aided Tran - slation)ವನ್ನು ಅಭಿವೃದ್ಧಿ ಪಡಿಸಿದೆ. ಈ MAT ವ್ಯವಸ್ಥೆ ಇಂಗ್ಲೀಷಿನಿಂದ ಕನ್ನಡಕ್ಕೆ ಪಠ್ಯವನ್ನು ಭಾಷಾಂತರ ಮಾಡಲು ರೂಪಿಸಿದ್ದಾಗಿದೆ. ಇದರಲ್ಲಿ ಕನ್ನಡ ಪದಕೋಶ, ಇಂಗ್ಲೀಷ್- ಕನ್ನಡ ದ್ವಿಭಾಷಾ ಪದಕೋಶ, ಒಂದು ರೀತಿಯ Thesarus ಮಾರ್ಫಾಲಜಿಕಲ್ ವಿಶ್ಲೇಷಕ ಮತ್ತು ಉತ್ಪಾದಕ (generatr) ಹಾಗೂ ಭಾಷಾಂತರಿಸಲು ಬೇಕಾದ ಕೆಲವು ಪರಿಸರಗಳು ಮತ್ತು

ಸಂಕಲನ ಸಂಪಾದನೆಗೆ ಬೇಕಾದ ಪರಿಕರಗಳನ್ನು ಹೊಂದಿದೆ. ಕನ್ನಡ ಪಾರ್ಸರ್‌ನ ನಮೂನೆಯನ್ನು ಅಭಿವೃದ್ಧಿ ಪಡಿಸಲಾಗಿದೆ. ಈ MAT ವ್ಯವಸ್ಥೆ(Universal Clause Structure Grammar (UCSG)ನ್ನು ಮತ್ತು ನಾನು ಅಭಿವೃದ್ಧಿ ಪಡಿಸಿದ ಸಂಬಂಧಿಸಿದ ವ್ಯವಸ್ಥೆಗಳನ್ನು ಆಧರಿಸಿದೆ. ಪಠ್ಯ ಪರಿಷ್ಕರಣೆಯ ಪರಿಕರಗಳನ್ನು ಅಭಿವೃದ್ಧಿ ಪಡಿಸಲಾಗಿದೆ. ಕೇಂದ್ರ ಸರಕಾರದ ಮಾಹಿತಿ ತಂತ್ರಜ್ಞಾನ ಸಚಿವಾಲಯದ ಹೊಸ ಯೋಜನೆಯಡಿ ಹೈದರಾಬಾದ್ ವಿಶ್ವವಿದ್ಯಾಲಯವು ತೆಲುಗು ಭಾಷೆಯನ್ನು ಮುಖ್ಯವಾಗಿಸಿ ಕೊಂಡು ಹಲವಾರು ವಿಷಯಗಳನ್ನು ಸಂಶೋಧನೆಗೆ ಎತ್ತಿಕೊಂಡಿದೆ. ಕನ್ನಡ ಮತ್ತು ತೆಲುಗು ಭಾಷೆಗಳು ರಚನೆಯ ದೃಷ್ಟಿಯಿಂದ ಬಹಳ ಸಾಮ್ಯತೆಯನ್ನು ಹೊಂದಿರುವುದರಿಂದ, ತೆಲುಗು ಭಾಷೆಯನ್ನು ಕುರಿತು ನಡೆದ ಬಹಳ ಕಾರ್ಯಗಳನ್ನು ಕನ್ನಡಭಾಷೆಗೂ ಅನ್ವಯಿಸಬಹುದು.

ಕಾನಪುರದ IIT (Indian Institute of Technology), ಒಂದು ಭಾರತೀಯ ಭಾಷೆಯಿಂದ ಮತ್ತೊಂದು ಭಾರತೀಯ ಭಾಷೆಗೆ ಭಾಷಾಂತರಿಸುವ ಅನುಸಾರಕ ವ್ಯವಸ್ಥೆಗಾಗಿ ಕನ್ನಡ-ಹಿಂದಿ ಪದಕೋಶವನ್ನು ಸಿದ್ಧಪಡಿಸಿದೆ.

ಇದಲ್ಲದೆ ಹಲವು ಖಾಸಗಿ ಸಂಸ್ಥೆಗಳೂ ಮತ್ತು ವ್ಯಕ್ತಿಗಳು ಪದ ಸಂಸ್ಕರಣೆ, ಅಕ್ಷರ ರೂಪಗಳಿಗೆ ಸಂಬಂಧಿಸಿದ ಹಲವಾರು ಪರಿಕರಗಳನ್ನು ಅಭಿವೃದ್ಧಿ ಪಡಿಸುವ ನಿಟ್ಟಿನಲ್ಲಿ ಕೆಲಸ ಮಾಡುತ್ತಿದ್ದಾರೆ. ಆದರೆ ಇದುವರೆಗೂ ನಡೆದಿರುವ ಕಾರ್ಯ ಅತ್ಯಲ್ಪ, ಆಗಬೇಕಾದದ್ದು ಬೆಟ್ಟದಷ್ಟಿದೆ.

೪. ಕ್ರಿಯಾ ಯೋಜನೆ : ಕೇವಲ ಒಬ್ಬ ವ್ಯಕ್ತಿಯಿಂದ ಅಥವಾ ಸಂಸ್ಥೆಯಿಂದ ನಿಗದಿತ ಸಮಯದಲ್ಲಿ ಅಗತ್ಯವಿರುವ ಕಾರ್ಯಗಳು ಆಗಲು ಸಾಧ್ಯವಿಲ್ಲ ಎನ್ನುವುದು ಸ್ಪಷ್ಟವಾಗಿದೆ. ನಮ್ಮ ಉದ್ದೇಶ ಬೇರೆಯವರು ತಲುಪಿರುವ ಹಂತವನ್ನು ಮುಟ್ಟುವುದಲ್ಲ. ಅದನ್ನು ಮೀರಿ ದಾರಿ ತೋರುವುದು, ಇದಕ್ಕೆ ಸಾಂಘಿಕ ಪ್ರಯತ್ನವೊಂದೇ ಪರಿಹಾರ, ಈಗಿನ ಅಗತ್ಯದ ಮೊದಲ ಹಂತವೇ ಹಲವಾರು ಸಂಸ್ಥೆಗಳು ಬೇರೆಬೇರೆಯಾಗಿ ಕೆಲಸ ಮಾಡುತ್ತಿರುವ ಸದಸ್ಯರನ್ನು ಒಗ್ಗೂಡಿಸಿ ತಂಡವಾಗಿ ರೂಪಿಸುವುದು. ಭಾಷಾ ಸಂಸ್ಕರಣೆಯ ಕಾರ್ಯ ಜಾಲವನ್ನು ಗಮನಿಸಿದರೆ ವಿವಿಧ ಶಿಸ್ತುಗಳಲ್ಲಿನ ವಿದ್ವಾಂಸರು ಒಟ್ಟಿಗೆ ಸೇರಿ, ಏಕೋದ್ದೇಶ ಕ್ಕಾಗಿ ಕೆಲಸಮಾಡಬೇಕಾಗಿರುವ ಅಗತ್ಯ ಮನದಟ್ಟಾಗುತ್ತದೆ. ಭಾಷಾತಜ್ಞರು, ಭಾಷಾವಿಜ್ಞಾನಿ ಗಳು, ಕಂಪ್ಯೂಟರ್ ವಿಜ್ಞಾನಿಗಳು, ಇಂಜಿನಿಯರ್‌ಗಳು, ಸಾಫ್ಟ್‌ವೇರ್ ಇಂಜಿನಿಯರ್‌ಗಳು, ಇಂಟರ್‌ನೆಟ್ ಮತ್ತು Web ಪರಿಣಿತರು, ಗಣಿತ ವಿದ್ವಾಂಸರು ಸಂಖ್ಯಾಶಾಸ್ತ್ರಜ್ಞರು ಇತ್ಯಾದಿ ಗಳೆಲ್ಲರ ಅಗತ್ಯವಿದೆ.

ಮುಖ್ಯ ಗುರಿ ಮೂಲಭೂತ ಮಾಹಿತಿಯನ್ನು ಸೃಷ್ಟಿಸುವುದಾಗಬೇಕಾಗುತ್ತದೆ. ಸಾದಾ, ಸೇರ್ಪಡಿತ, ಲಕ್ಷಣ ನಿರ್ವಚಿತ ಮತ್ತು ಮಾತಿನ ಕಾರ್ಪಸ್‌ಗಳನ್ನು (ಪದಸಮುಚ್ಛಯಗಳನ್ನು), KWIC ಸೂಚಿಗಳನ್ನು, ಅಂಕಿಸಂಖ್ಯಾ ವಿಶ್ಲೇಷಣೆ ಇತ್ಯಾದಿಗಳನ್ನು ಸೃಷ್ಟಿಸುವುದು ಮೊಟ್ಟ

ಮೊದಲಿಗೆ ಆಗಬೇಕಾದ ಕೆಲಸ. ಇವುಗಳು ಬಹಳ ಕ್ಲಿಷ್ಟವಾದದ್ದು, ಸಮಯಬೇಡು ವಂಥದ್ದು ಮತ್ತು ಸವಾಲೆಸೆಯುವಂಥದ್ದಾಗಿರುವುದರಿಂದ, ಲಭ್ಯವಿರುವ ಉತ್ತಮ ಸಂಪನ್ಮೂಲವನ್ನು ಬಳಸಬೇಕಾಗುತ್ತದೆ.

ಪದಕೋಶ, ಮಾರ್ಫಾಲಜಿ, ಕಾಗುಣಿತ ಪರಿಶೀಲಕ, ಲಕ್ಷಣ ನಿರ್ವಚನ ಮತ್ತು ಭಾಷಾಂತರ ಇತ್ಯಾದಿ ಕಾರ್ಯಗಳನ್ನು ಭಾಷಾಶಾಸ್ತ್ರ ಮತ್ತು ಕಂಪ್ಯೂಟರ್ ತಂತ್ರಜ್ಞಾನ ಎರಡು ಇರುವ ಸಂಸ್ಥೆಗಳು ಕೈಗೊಳ್ಳಬೇಕಾಗುತ್ತದೆ.

OCR, ಮಾತಿನ ಪರಿಷ್ಕರಣ ಇತ್ಯಾದಿ ವಿಶೇಷ ಕೆಲಸಗಳನ್ನು ಅದರಲ್ಲಿ ಬಳಕೆ ವಲಯಗಳ ಕಾರ್ಯವನ್ನು ವ್ಯಕ್ತಿಗತವಾಗಿ ಅಥವಾ ಸಾಂಸ್ಥಿಕ ಮಟ್ಟದಲ್ಲಿ ಆ ವಿಷಯದಲ್ಲಿ ಪರಿಣಿತವಾಗಿರುವ ಸಂಸ್ಥೆಗಳೇ ಮಾಡಬೇಕಾಗುತ್ತದೆ. ಆಸಕ್ತಿ ಮತ್ತು ಹಿನ್ನೆಲೆ ಇರುವವರು ಮಾಡಬೇಕಾಗುತ್ತದೆ. ಸಹಕಾರ ಮತ್ತು ಒಟ್ಟುಗೂಡಿವಿಕೆ ಹಲವಾರು ರೂಪಗಳನ್ನು ಧರಿಸಬಹುದು. ವಿದ್ಯಾರ್ಥಿ ಸಂಶೋಧನಾ ಯೋಜನೆ, Ph.D. ಯೋಜನೆಗಳು, ಯೋಜನೆಯ ಹಂತದಲ್ಲಿ ಸಂಸ್ಥೆಗಳ ಒಟ್ಟುಗೂಡುವಿಕೆ, ವ್ಯಕ್ತಿಗಳ ಒಟ್ಟುಗೂಡುವಿಕೆ ಇತ್ಯಾದಿ. ಹೈದರಾಬಾದ್ ವಿಶ್ವವಿದ್ಯಾಲಯವು ಎಲ್ಲ ಆಸಕ್ತ ವ್ಯಕ್ತಿ ಮತ್ತು ಸಂಸ್ಥೆಗಳೊಂದಿಗೆ ಎಲ್ಲ ಹಂತಗಳಲ್ಲೂ ಭಾಗವಹಿಸಲು, ಸಹಕರಿಸಲು, ಸಂಚಾಲನೆಯಲ್ಲಿ ಭಾಗವಹಿಸಲು ಸಿದ್ಧವಿದೆ.

ಉಲ್ಲೇಖಗಳು

1. Bran Boguraev, Ted Briscoe (Eds), Computational Lexicography for Natural Langauage Processing, Longaman. 1989.
2. Eugene Charniak, "Statistical Language Learning", MIT Press. 1993.
3. Gerald Salton, "Automatic Text Processing" , Addison-Wesley. 1989.
4. K.Narayana Murthy, "Electronic Dictionaries and Computational Tools", Linguistics Today, Vol 1, No 1, July 1997 pp 34-50.
5. K.Narayana Murthy, A.Sivasankara Reddy "Universal Clasue Structure Grammar" Computer Science and Informatics, Vol 27, No 1, March 1997 Special Issue on Natural Langauage Processing and Machine Learning, pp 26-38.
6. K. Narayana Murthy, "Universal Clause Structure Grammar and the Syntax of Relatively Free Word Order Languages", South Asian Language Review, Vol VII, No 1, Jan 1997, pp 47-64.
7. K.Narayana Murthy, "MAT:A Machine Assisted Translation System", Proceedings of the Fifth Natural Language Pacific Rim Symposium, NLPRS-99, 5-7 November, Beijing, China.
8. K. Narayana Murthy, "A Network and Process Model for Morphological Analysis/Generation", To appear in the Proceedings of the Second International Conference on South Asian Languages, 9-11 January 1999, Punjabi University, Patiala, India.

9. K. Narayana Murthy "An Indexing Technique for Efficient Retrieval from Large Dictionaries" To appear in the Proceedings of National Conference on Information Technology NCIT-97, 21-23 December 1997, Bhubaneswar.
10. K. Narayana Murthy "Parsing Telugu in the UCSG Formalism" Proceedings of the Indian Congress on Knowledge and Language, Vol 2, pp 1-16 11-17 January 1996, Central Institute of Indian Languages, Mysore.
11. K. Vasuprada, K.Narayana Murthy, "Part-of-Speech Tagging using a Tritag Model", To appear in the Proceedings of the Second National Symposium on Quantitative Linguistics, 28-29 February 2000. Indian Statistical Institute. Culcutta.
12. Maric-Francine Moens, "Automatic Indexing and Abstracting of Document Texts", Kluwer Academic Publishers, 2000.
13. Michael P. Oakes, "Statistics for Corpus Linguistics", Edinburgh University Press, 1998.
14. S.N.Srihari, "Computer Text Recognition and Error Correction", IEEE Computer Society Press, 1984.