

**W3C Workshop**

**The Multilingual Web – The Way Ahead**

Discussion Topic

# **Speech Technologies for the Multilingual Web**



# The web is becoming more and more VOCAL!

Speech or speech-related contents are increasingly being brought onto the Internet in a variety of forms:

- e-learning classes (e.g., Khan Academy, language classes, NLP courses, etc.)
- motivational multi-purpose talks (e.g., TED talks)
- lyrics on music videos
- movie subtitles
- broadcast news captioning
- advertising
- etc.



# State-of-the-Art Core Speech Technologies

- speech understanding
- summarization of spoken documents
- classification, question-answering, and rich transcription of multimedia documents
- computer assisted language learning
- speech-to-speech machine translation



# Socio-Economic Importance of Spoken Language

- Complements the more traditional document-content based web
- Further strengthens and make more appealing and dynamic the information access over the internet
- Creates new opportunities to spread and share knowledge on a global scale
- Helps businesses advertise, market and sell their products and services
- Addresses educational inclusion and special communication needs
- Extremely popular among younger generations
- Enriches the social value of the web
- Facilitates human communication



# Conclusion

Speech Technology **CANNOT** be absent  
from the multilingual web discussion!

