* Web mining ສາມາດແບ່ງອອກເປັນສາມປະເພດຄື:

1. web content mining ແມ່ນຄໍາຮ້ອງສະຫມັກຂອງການສະກັດຂໍ້ມູນທີ່ເປັນປະໂຫຍດຈາກເນື້ອໃນຂອງເອກະສານເວັບໄຊຕ໌.

* ການຂຸດຄົ້ນເນື້ອຫາເວັບເປັນພາສາຕ່າງປະເທດ
* ຈີນ

ລະຫັດພາສາຂອງຄໍາສັບພາສາຈີນແມ່ນສັບສົນຫຼາຍເມື່ອທຽບກັບພາສາອັງກິດ.

1. Web usage miningແມ່ນການນໍາໃຊ້ເຕັກນິກການຂຸດຄົ້ນຂໍ້ມູນເພື່ອຄົ້ນພົບຮູບແບບການນໍາໃຊ້ຈາກຂໍ້ມູນເວັບ, ເພື່ອເຂົ້າໃຈແລະໃຫ້ບໍລິການຄວາມຕ້ອງການຂອງແອັບພລິເຄຊັນເວັບ.

* **Web server data**: ບັນທຶກຜູ້ໃຊ້ຖືກເກັບລວບລວມໂດຍເຄື່ອງແມ່ຂ່າຍເວັບ.
* **Application server data**:ເຄື່ອງແມ່ຂ່າຍຂອງຄໍາຮ້ອງສະຫມັກການຄ້າມີລັກສະນະທີ່ສໍາຄັນເພື່ອເຮັດໃຫ້ຄໍາຮ້ອງສະຫມັກອີຄອມເມີຊທີ່ສ້າງຂຶ້ນຢູ່ເທິງສຸດຂອງພວກເຂົາດ້ວຍຄວາມພະຍາຍາມຫນ້ອຍ.
* **Application level data**:ປະເພດຂອງເຫດການ ໃໝ່ ສາມາດຖືກ ກຳ ນົດໃນແອັບພລິເຄຊັນ, ແລະການບັນທຶກສາມາດເປີດໃຫ້ພວກເຂົາໄດ້, ສະນັ້ນການສ້າງປະຫວັດສາດຂອງເຫດການທີ່ ກຳ ນົດພິເສດເຫຼົ່ານີ້.
* Pros Web usage mining

ການຂຸດຄົ້ນການນໍາໃຊ້ເວັບມີຄວາມໄດ້ປຽບຫຼາຍທີ່ເຮັດໃຫ້ເຕັກໂນໂລຢີນີ້ດຶງດູດເອົາບໍລິສັດລວມທັງອົງການຂອງລັດຖະບານ. ເທກໂນໂລຍີນີ້ໄດ້ຊ່ວຍໃຫ້ອີຄອມເມີຊສາມາດເຮັດການຕະຫຼາດສ່ວນບຸກຄົນ, ເຊິ່ງໃນທີ່ສຸດກໍ່ເຮັດໃຫ້ປະລິມານການຄ້າສູງຂຶ້ນ. ອົງການຂອງລັດຖະບານກໍາລັງໃຊ້ເຕັກໂນໂລຢີນີ້ເພື່ອຈັດປະເພດໄພຂົ່ມຂູ່ແລະການຕໍ່ສູ້ກັບການກໍ່ການຮ້າຍ.

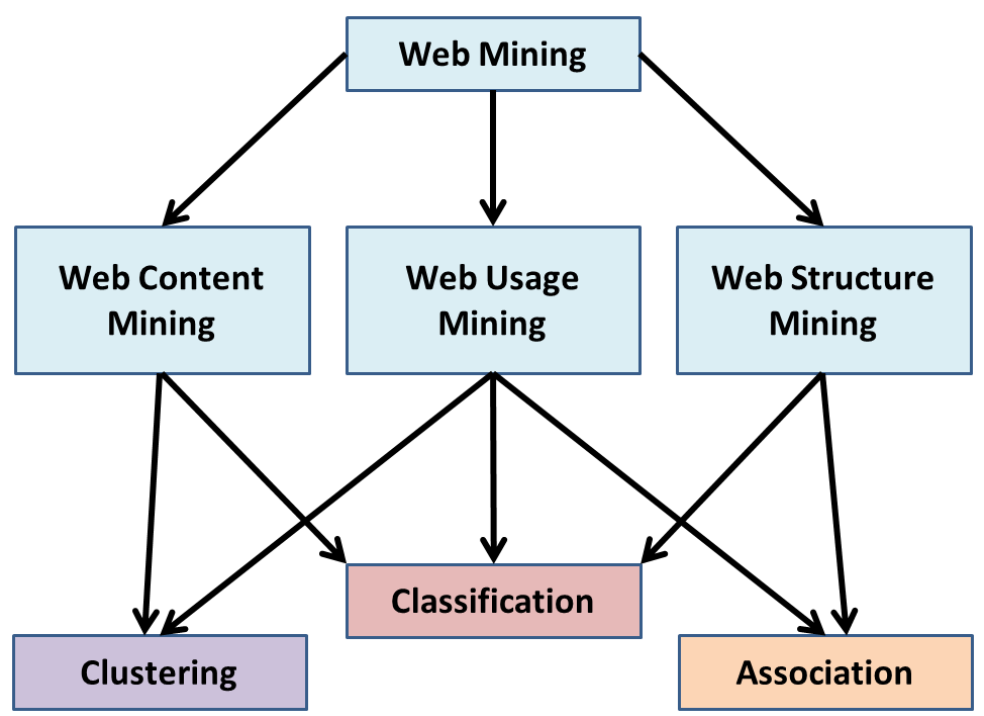
* Cons Web usage mining

ການຂຸດຄົ້ນການນໍາໃຊ້ເວັບໂດຍຕົວມັນເອງບໍ່ໄດ້ສ້າງບັນຫາ, ແຕ່ເຕັກໂນໂລຢີນີ້ເມື່ອນໍາໃຊ້ກັບຂໍ້ມູນສ່ວນຕົວອາດຈະເຮັດໃຫ້ເກີດຄວາມກັງວົນ. ບັນຫາດ້ານຈັນຍາບັນທີ່ຖືກວິພາກວິຈານທີ່ສຸດທີ່ກ່ຽວຂ້ອງກັບການຂຸດຄົ້ນການນໍາໃຊ້ເວັບໄຊຕ໌ແມ່ນການບຸກລຸກຂອງຄວາມເປັນສ່ວນຕົວ.

1. Web structure mining ຄໍາຮ້ອງສະຫມັກຂອງການຄົ້ນພົບຂໍ້ມູນໂຄງສ້າງຈາກເວັບໄຊຕ໌.

* ຄໍາສັບກ່ຽວກັບໂຄງສ້າງເວັບ:
* ກຣາຟເວັບ: ເສັ້ນກຣາບຊີ້ບອກທີ່ເປັນຕົວແທນຂອງເວັບ.
* Node: ຫນ້າເວັບໃນກາຟ.
* ຂອບ: hyperlinks.
* ໃນລະດັບ: ຈໍານວນຂອງການເຊື່ອມຕໍ່ຊີ້ໄປຫາ node ໂດຍສະເພາະ.
* Out degree: ຈໍານວນຂອງການເຊື່ອມຕໍ່ທີ່ສ້າງຂຶ້ນຈາກ node ໂດຍສະເພາະ.
* Classification ແມ່ນບັນຫາໃນການກໍານົດວ່າຊຸດຂອງປະເພດ (ປະຊາກອນຍ່ອຍ) ການສັງເກດການ (ຫຼືການສັງເກດການ) ເປັນຂອງ.

ຕົວຢ່າງແມ່ນການມອບຫມາຍອີເມລ໌ໃຫ້ກັບຫ້ອງຮຽນ "spam" ຫຼື "ບໍ່ແມ່ນ spam", ແລະມອບຫມາຍການບົ່ງມະຕິໃຫ້ຄົນເຈັບໂດຍອີງຕາມຄຸນລັກສະນະທີ່ສັງເກດເຫັນຂອງຄົນເຈັບ (ເພດ, ຄວາມດັນເລືອດ, ມີຫຼືບໍ່ມີອາການບາງຢ່າງ, ແລະອື່ນໆ). .

* Clustering ແມ່ນໜ້າທີ່ຂອງການຈັດກຸ່ມຂອງວັດຖຸໃນລັກສະນະທີ່ວັດຖຸຢູ່ໃນກຸ່ມດຽວກັນ (ເອີ້ນວ່າກຸ່ມ) ມີຄວາມຄ້າຍຄືກັນຫຼາຍກວ່າ (ໃນບາງຄວາມໝາຍ) ຕໍ່ກັນຫຼາຍກວ່າກຸ່ມອື່ນ (ກຸ່ມ).
* Association ແມ່ນ ວິທີ [ການຮຽນຮູ້ເຄື່ອງຈັກໂດຍອີງໃສ່ກົດລະບຽບ](https://en-m-wikipedia-org.translate.goog/wiki/Rule-based_machine_learning?_x_tr_sl=en&_x_tr_tl=lo&_x_tr_hl=lo&_x_tr_pto=wapp) ສໍາລັບການຄົ້ນພົບການພົວພັນທີ່ຫນ້າສົນໃຈລະຫວ່າງຕົວແປໃນຖານຂໍ້ມູນຂະຫນາດໃຫຍ່ 
* **Web analytics technologies**:ເທກໂນໂລຍີການວິເຄາະເວັບ
* ມີຢ່າງຫນ້ອຍສອງປະເພດຂອງການວິເຄາະເວັບ, ການວິເຄາະເວັບນອກແລະໃນເວັບໄຊທ໌.

1.Off-site analytics ຫມາຍເຖິງການວັດແທກແລະການວິເຄາະເວັບໂດຍບໍ່ຄໍານຶງເຖິງວ່າບຸກຄົນໃດຫນຶ່ງເປັນເຈົ້າຂອງຫຼືຮັກສາເວັບໄຊທ໌. ມັນປະກອບມີການວັດແທກຜູ້ຊົມທີ່ມີທ່າແຮງຂອງເວັບໄຊທ໌ (ໂອກາດ), ການແບ່ງປັນສຽງ (ການເບິ່ງເຫັນ), ແລະ buzz (ຄໍາເຫັນ) ທີ່ເກີດຂຶ້ນໃນອິນເຕີເນັດທັງຫມົດ.

2.On-site analytics , ທົ່ວໄປຫຼາຍຂອງສອງ, ການວັດແທກພຶດຕິກໍາຂອງຜູ້ເຂົ້າຊົມຫນຶ່ງຄັ້ງໃນເວັບໄຊທ໌ສະເພາະໃດຫນຶ່ງ.ນີ້ປະກອບມີໄດເວີແລະການແປງຂອງມັນ; ຕົວຢ່າງ, ລະດັບຂອງຫນ້າທີ່ດິນທີ່ແຕກຕ່າງກັນທີ່ກ່ຽວຂ້ອງກັບການຊື້ອອນໄລນ໌. ການວິເຄາະເວັບໄຊທ໌ຢູ່ໃນເວັບໄຊທ໌ວັດແທກການປະຕິບັດຂອງເວັບໄຊທ໌ສະເພາະໃດຫນຶ່ງໃນສະພາບການການຄ້າ. ຂໍ້ມູນນີ້ໂດຍທົ່ວໄປແລ້ວຈະຖືກປຽບທຽບກັບຕົວຊີ້ວັດການປະຕິບັດທີ່ສໍາຄັນສໍາລັບການປະຕິບັດແລະຖືກນໍາໃຊ້ເພື່ອປັບປຸງເວັບໄຊທ໌ຫຼືການໂຄສະນາການຕອບສະຫນອງຂອງຜູ້ຊົມ. Google Analytics ແລະ Adobe Analytics ແມ່ນການບໍລິການການວິເຄາະເວັບໄຊທ໌ທີ່ຖືກນໍາໃຊ້ຢ່າງກວ້າງຂວາງທີ່ສຸດ; ເຖິງແມ່ນວ່າເຄື່ອງມືໃຫມ່ກໍາລັງເກີດຂື້ນທີ່ສະຫນອງຂໍ້ມູນຊັ້ນເພີ່ມເຕີມ, ລວມທັງແຜນທີ່ຄວາມຮ້ອນແລະການຫຼີ້ນໃຫມ່ຂອງເຊດຊັນ.

* **Web analytics data sources:ແຫຼ່ງຂໍ້ມູນການວິເຄາະເວັບ**
* ຂໍ້ມູນສ່ວນໃຫຍ່ແມ່ນມາຈາກສີ່ແຫຼ່ງ:

1 ຂໍ້ມູນການຮ້ອງຂໍ HTTP ໂດຍກົງ: ມາຈາກຂໍ້ຄວາມການຮ້ອງຂໍ HTTP ໂດຍກົງ (ສ່ວນຫົວຄໍາຮ້ອງຂໍ HTTP).

2 ຂໍ້ມູນລະດັບເຄືອຂ່າຍແລະເຊີຟເວີທີ່ສ້າງຂຶ້ນທີ່ກ່ຽວຂ້ອງກັບຄໍາຮ້ອງຂໍ HTTP: ບໍ່ແມ່ນສ່ວນຫນຶ່ງຂອງຄໍາຮ້ອງຂໍ HTTP, ແຕ່ມັນຈໍາເປັນຕ້ອງມີການສົ່ງຄໍາຮ້ອງຂໍທີ່ປະສົບຜົນສໍາເລັດ - ຕົວຢ່າງ, ທີ່ຢູ່ IP ຂອງຜູ້ຮ້ອງຂໍ.

3 ຂໍ້ມູນລະດັບແອັບພລິເຄຊັນທີ່ສົ່ງກັບຄໍາຮ້ອງຂໍ HTTP: ສ້າງຂຶ້ນ ແລະປະມວນຜົນໂດຍໂປຣແກຣມລະດັບແອັບພລິເຄຊັນ (ເຊັ່ນ: JavaScript, PHP, ແລະ ASP.Net), ລວມທັງເຊດຊັນ ແລະການສົ່ງຕໍ່. ປົກກະຕິແລ້ວສິ່ງເຫຼົ່ານີ້ຖືກຈັບໂດຍບັນທຶກພາຍໃນແທນທີ່ຈະເປັນການບໍລິການການວິເຄາະເວັບສາທາລະນະ.

4 ຂໍ້ມູນພາຍນອກ: ສາມາດຖືກລວມເຂົ້າກັບຂໍ້ມູນຢູ່ໃນເວັບໄຊທ໌ເພື່ອຊ່ວຍເພີ່ມຂໍ້ມູນພຶດຕິກໍາຂອງເວັບໄຊທ໌ທີ່ອະທິບາຍຂ້າງເທິງແລະຕີຄວາມຫມາຍການນໍາໃຊ້ເວັບ. ຕົວຢ່າງ, ທີ່ຢູ່ IP ປົກກະຕິແລ້ວແມ່ນກ່ຽວຂ້ອງກັບພາກພື້ນທາງພູມສາດແລະຜູ້ໃຫ້ບໍລິການອິນເຕີເນັດ, ອັດຕາການເປີດອີເມວແລະການຄລິກຜ່ານ, ຂໍ້ມູນການໂຄສະນາທາງອີເມວ, ການຂາຍ, ແລະປະຫວັດການນໍາ, ຫຼືປະເພດຂໍ້ມູນອື່ນໆຕາມຄວາມຕ້ອງການ.

* **Web server log file analysis:ການວິເຄາະໄຟລ໌ບັນທຶກຂອງເຄື່ອງແມ່ຂ່າຍເວັບ**

ເຊີບເວີເວັບບັນທຶກບາງທຸລະກໍາຂອງເຂົາເຈົ້າຢູ່ໃນໄຟລ໌ບັນທຶກ. ມັນໄດ້ຖືກຮັບຮູ້ທັນທີວ່າໄຟລ໌ບັນທຶກເຫຼົ່ານີ້ສາມາດອ່ານໄດ້ໂດຍໂຄງການເພື່ອສະຫນອງຂໍ້ມູນຄວາມນິຍົມຂອງເວັບໄຊທ໌